

DWH MODERNISIERUNG

EIN PRAXISBERICHT

„Das DWH ist tot, lang lebe das DWH“ – was denn nun? Die meisten Unternehmen und Organisationen beschäftigen sich mit der Frage, ob ihr DWH noch zeitgemäß ist oder ob es nicht durch Hadoop, die Cloud und andere neue Technologien bereits obsolet geworden ist. Wir wollen der Frage mit einem Bericht aus der Beraterpraxis nachgehen.

EINLEITUNG

Gründe für den Wunsch nach einer Modernisierung des DWH gibt es viele: hohe Kosten für Entwicklung und Betrieb, Unzufriedenheit der Nutzer, neue Technologien, die größeren Nutzen bei geringeren Kosten versprechen.

Welchen der Versprechungen kann man Glauben schenken? Welche Einwände und Bedenken muss man ernst nehmen? Und vor allem: wann ist es wirklich an der Zeit, eine umfassende Modernisierung der Data-Warehouse-Landschaft in Angriff zu nehmen?

DEFINITION

Bevor wir jedoch auf die möglichen Beweggründe für eine Modernisierung sowie die daraus abzuleitenden Handlungsoptionen eingehen, muss zuerst einmal eine Begriffsbestimmung erfolgen: wann handelt es sich tatsächlich um eine DWH-Modernisierung und wie kann diese gegenüber geringfügigen Änderungen wie einem Datenbank-Upgrade abgegrenzt werden?

Folgt man gängigen Definitionen kann Modernisierung auf vielfältige Art und Weise interpretiert werden. Der Begriff steht sowohl für einfache Softwareupdates als auch für die komplette Neuausrichtung der DWH-Architektur. Zudem zählen neue modellgetriebene Entwicklungsmethoden und agile Projektführungsansätze auch zum Modernisierungsbegriff.

Aus unserer Sicht handelt es sich dabei um alle Maßnahmen, die es einer Organisation ermöglichen, ihre Data Warehouse Umgebung relevant und kompetitiv zu halten und an neue geschäftliche und technologische Anforderungen auszurichten. Einfache Software-Updates, die keine neuen Anforderungen bedienen, wären nicht ausreichend um schon als Modernisierung gewertet zu werden.

URSACHEN

Ebenso wie kein Arzt ohne Anamnese eine Diagnose stellen und einen Therapieversuch machen würde, muss man sich auch bei der Modernisierung des Data Warehouse zuerst mit den Symptomen auseinandersetzen: wenn bspw. das DWH nicht die für Analysen benötigten Daten enthält, so

wird auch eine Migration auf eine neue Plattform nicht zur Steigerung der Nutzerzufriedenheit beitragen.

Ohne Anspruch auf theoretische Vollständigkeit zeigt die Praxis, dass zwischen den folgenden Ursachen unterschieden werden muss, auch wenn in „schweren Fällen“ mehrere, wenn nicht gar alle vorliegen können:



MANGELNDE LEISTUNG

Wenn sich Nutzer über lange Antwortzeiten beschweren, so wird dies natürlich oft an einer veralteten oder unterdimensionierten Plattform liegen, was vergleichsweise einfach durch ein Hardware-Upgrade behoben werden kann. Häufig jedoch sind derartige Probleme auf Schwächen der Architektur zurückzuführen, welche umfangreichere Sanierungsmaßnahmen erfordern.



FACHLICHE ABDECKUNG

Eine weitere Ursache für den Wunsch nach Modernisierung und auch das Entstehen der gefürchteten Schatten-IT in den Fachbereichen ist das Fehlen wesentlichen Datenbereiche im DWH. Während ein zusätzliches Quellsystem oft vergleichsweise einfach angebunden werden kann, können größere Defizite Änderungen am Datenmodell und den ETL-Strecken erfordern, die de facto einem Neubau gleichkommen.



FEHLENDES VERTRAUEN

Wenn man Fachbereiche nach den Gründen für ihre Unzufriedenheit oder gar die Nichtnutzung des DWH befragt, so werden oft unzureichende Datenqualität oder überhaupt fehlendes Wissen um den Datenhaushalt des DWH und dessen Zuverlässigkeit angeführt. Die Praxis zeigt, dass entsprechende Dokumentation wie z.B. in Form eines fachlichen Datenmodells oder Glossars oft deutlich schneller und mit weniger Aufwand zum Erfolg führen als technische Änderungen.



ANALYTISCHE MÖGLICHKEITEN

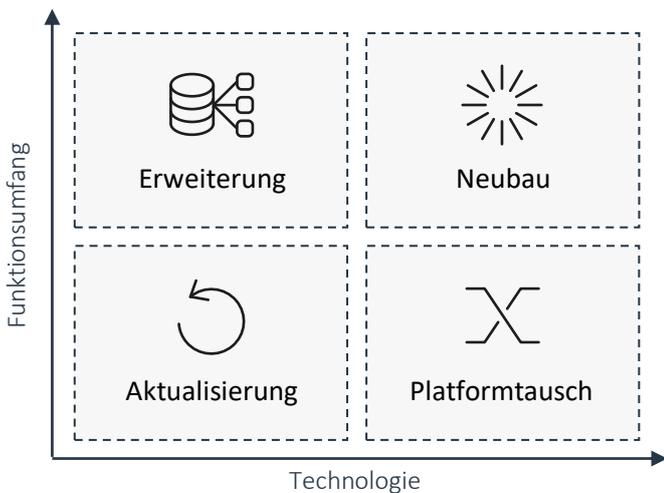
Nicht selten zeigt sich, dass der Wunsch nach Modernisierung von Unzulänglichkeiten der im Einsatz befindlichen Auswertungs- und Berichtswerkzeuge getrieben ist. Sofern dies tatsächlich als einzige oder zumindest massiv überwiegende Ursache identifiziert werden kann, ist eine Sanierung oft durch ein neues BI-Tool und eine Neugestaltung der Berichtslandschaft ohne nennenswerte Änderungen am DWH selbst möglich.

Wie eingangs erwähnt, ist im Alltag öfter eine Gemengelage an Ursachen anzutreffen. Trotzdem kann die o.a. Gliederung eine erste Orientierung zu den in Frage kommenden Maßnahmen bieten.

MODERNISIERUNGSSTRATEGIEN

Sobald Klarheit und allgemeine Übereinstimmung herrschen, worin die Ursachen liegen, gilt es, sich für eine der zur Verfügung stehenden Modernisierungsstrategien zu entscheiden. Auch wenn manchmal radikale Schritte notwendig sein können, so sei doch vor „falschen Propheten“ (oder auch Evangelisten) gewarnt. Weder ist das DWH tot noch ist Hadoop die alleinseligmachende Lösung.

Allgemein ist zwischen zwei Dimensionen der Modernisierung, nämlich der Technologie und dem Funktionsumfang zu unterscheiden:



AKTUALISIERUNG

Minimalinvasive Modernisierung durch Einsatz einer neuen Datenbankversion, um bspw. neue Funktionalitäten nutzen zu können oder auch eine geringfügige Änderung des Datenmodells oder anderer Komponenten zur Leistungssteigerung.

ERWEITERUNG

Nicht selten genügt die vorhandene Technologie den Ansprüchen, jedoch ist der Datenhaushalt und die darauf aufbauenden Berechnungen zu gering, um den Fachbereichen vernünftige Analysen und Berichte zu ermöglichen.

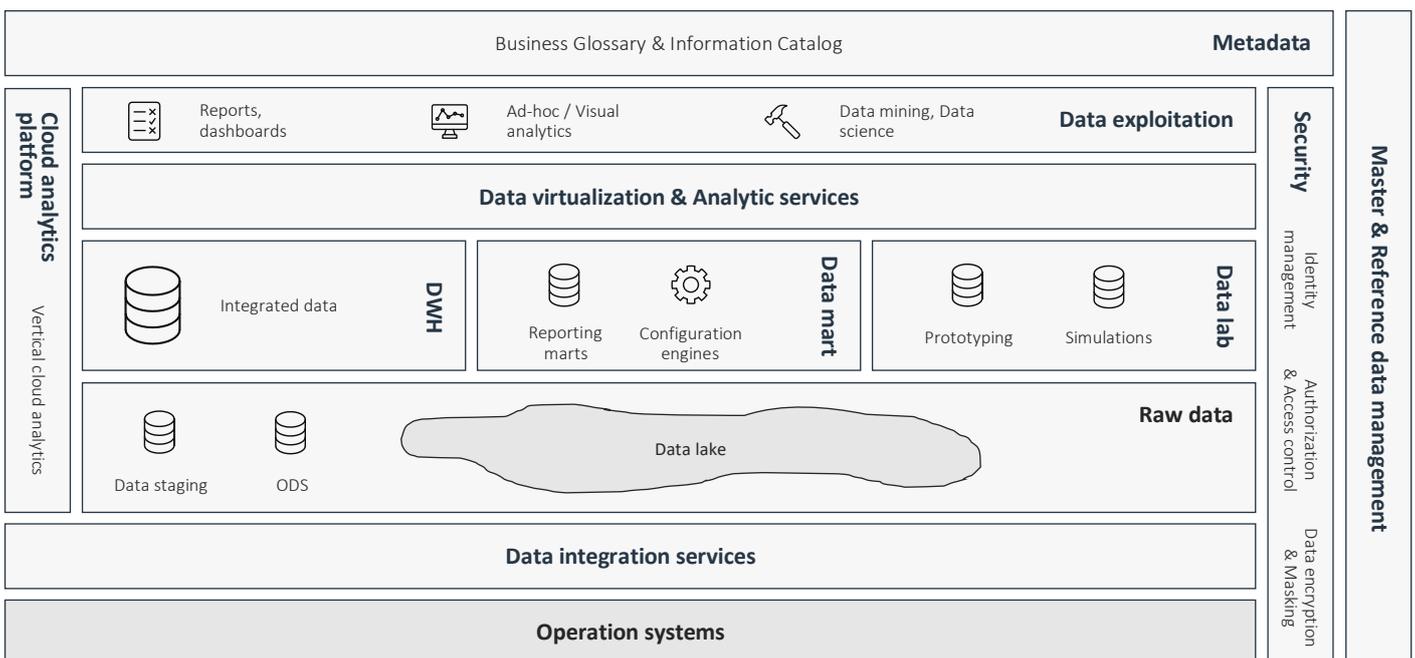
PLATTFORMTAUSCH

Der gegenläufige Fall zur Erweiterung, wo zwar der Datenhaushalt quantitativ und qualitativ den Ansprüchen im Wesentlichen genügt, jedoch die im Einsatz befindliche Technologie das Ende ihrer Lebensdauer erreicht hat.

NEUBAU

Die radikalste Form der Erneuerung, oft auch als disruptiv bezeichnet. Hierbei wird neben einem Wechsel der Technologie auch der Datenhaushalt bzw. Leistungsumfang ausgebaut. Häufig gibt eine der beiden Dimensionen den Ausschlag, auch die andere in Angriff zu nehmen.

Wie schon bei den Ursachen, so gilt auch hier, dass in der Praxis gewisse Mischformen anzutreffen sind. Als besonders erfolgversprechend haben sich dabei Ansätze erwiesen, die eine „friedliche Koexistenz“ verschiedener Technologien und Methodiken für die unterschiedlichen Einsatzzwecke nicht nur zulassen, sondern aktiv fördern. Für nähere Informationen dazu möchten wir auf einen kürzlich erschienenen Artikel zum Thema logisches DWH und multimodale Governance verweisen.



Es geht dabei darum, die verschiedenen Konzepte wie klassisches DWH, Data Lake und Cloud bzw. andere Formen der Virtualisierung nicht als konkurrierende, sondern vielmehr komplementäre Lösungskomponenten zu sehen, wobei die fachlichen Metadaten (z.B. ein Glossar oder der von uns vertretene Ansatz des fachlichen Datenmodells) als logische Klammer dienen, die den Zusammenhalt der einzelnen Bestandteile gewährleistet.

HERAUSFORDERUNGEN UND RISIKEN

Vergleichbar dem Verschweigen von Nebenwirkungen eines verordneten Medikaments wäre es höchst fahrlässig, nicht auf die Herausforderungen und Risiken solcher Modernisierungsvorhaben einzugehen. Natürlich treffen die meisten auf Großprojekte im Allgemeinen zu, jedoch wollen wir die DWH-spezifischen Aspekte hervorheben.



FACHBEREICHSUNTERSTÜTZUNG

Auch wenn Modernisierungsvorhaben nicht selten von einer wohlmeinenden IT-Organisation in die Wege geleitet werden, so darf doch nicht auf die umfassende und frühzeitige Einbindung der Fachbereiche vergessen werden. Selbst wenn, was keinesfalls garantiert ist, die Probleme richtig erkannt werden, so hängt die Akzeptanz der angestrebten Lösung doch maßgeblich von der Einbindung der „Kunden“ ab. Dieser Aspekt gewinnt besonders an Bedeutung, wenn es darum geht, Fachbereichslösungen („Schatten-IT“) wieder in die Kontrolle der IT-Abteilung zu überführen.



ANFORDERUNGSERHEBUNG

Wie im Kapitel Ursachen dargestellt, sind umfangreiche Modernisierungsmaßnahmen vor allem dann erforderlich, wenn den Fachbereichen nicht die notwendigen Daten zur Verfügung stehen. Wer aus falsch verstandener Agilität gleich „lossprintet“, ohne die Datenanforderungen und analytischen Bedürfnisse sorg-

fältig erhoben und wirklich verstanden zu haben, wird oft in einer Sackgasse landen. Dies ändert jedoch nichts an dem Umstand, dass mehrjährige, monolithische Großprojekte heute – zurecht – keine Akzeptanz mehr finden. Unser Ansatz zur fachlichen Datenmodellierung bietet einen Ausweg aus diesem Dilemma.



GROßE WETTEN

Auch wenn die Versprechungen verlockend sind, so sei doch davor gewarnt, technologisch alles „auf eine Karte“ zu setzen. Die meisten Organisationen haben vielfältige und oft divergierende analytische Bedürfnisse; nur selten ist eine Technologie die richtige Antwort für alle diese Fragen. Eher geht es darum, ein vernünftiges Miteinander zu orchestrieren. Was darüber oft auch vernachlässigt wird, ist das fehlende Erfahrung im Umgang mit den neuen Technologien. Dies hat schon oft dazu geführt, dass neue Hardware im wahrsten Sinne des Wortes zum „Staubfänger“ wird, da die notwendigen Ressourcen weder in der Organisation noch am Arbeitsmarkt verfügbar sind.

ERFOLGSFAKTOREN

Wenn man nun den Entschluss zur DWH-Modernisierung gefasst hat, wie soll man ein derartiges Vorhaben in Angriff nehmen? Trotz zahlreicher erfolgreicher Projekte ist es uns noch nicht gelungen, ein einzelnes Wundermittel zu identifizieren. Jedoch gibt es zahlreiche Lösungskomponenten, die sich immer wieder und in den unterschiedlichsten Kundensituationen bewährt haben.

DATENGETRIEBENE PROJEKTPLANUNG

Trotz anfänglicher Skepsis zeigt sich oft, dass die Schnittmenge der Datenanforderungen zwischen den Fachbereichen viel größer ist als anfänglich angenommen bzw. eingestanden. Während Unterschiede in den Kennzahlen und Metriken in der Natur der Sache liegen, so werden zumeist doch die gleichen Basisdaten für deren Berechnung benötigt.

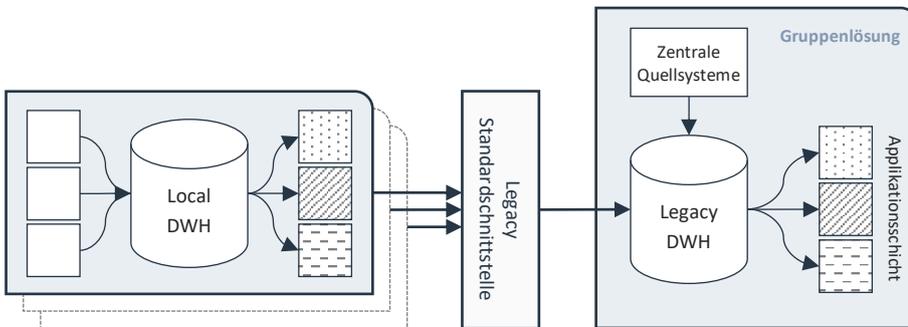
Einer der unleugbaren Vorteile von fachlichen Datenmodellen und anderen formalisierten Methoden zur Erhebung der Datenanforderungen ist eine detaillierte Sicht auf die Überschneidungen zwischen den Fachbereichen, welche ein unschätzbar wertvolles Instrument der Projektplanung darstellen kann.

In der Praxis nimmt dies in der Regel eine der folgenden Ausformungen an: entweder werden Datenbereiche nach dem Grad der Überschneidung zwischen den Fachbereichen für die Datenintegration priorisiert, oder es kann ein Datenbereich, der ausschließlich für eine kleine Anwendergruppe von Interesse ist, für Pilotzwecke identifiziert werden.

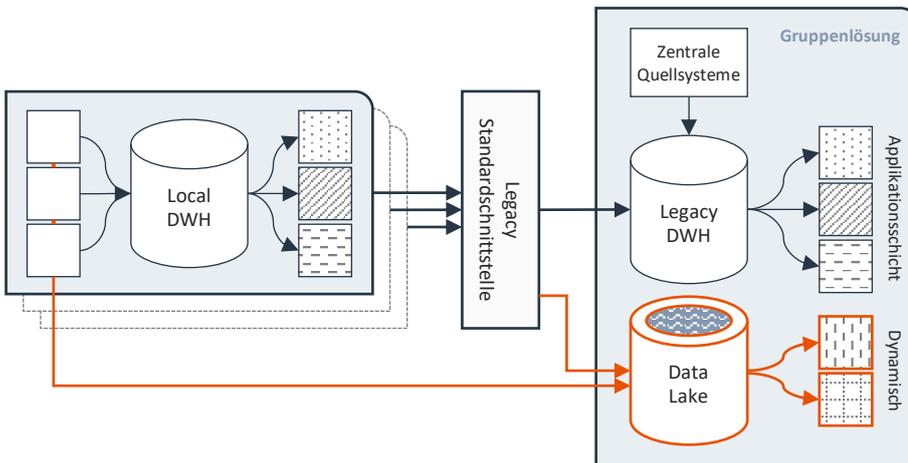
ÜBERGANGSARCHITEKTUREN

Während Ist- und Zielarchitektur für gewöhnlich ausreichend geplant und dokumentiert werden, erhalten die notwendigen Zwischenschritte vom einen zum anderen oft zu wenig Aufmerksamkeit. Gerade in großen und komplexen Organisationen, die z.B. auch die Bedürfnisse von Tochtergesellschaften oder Auslandsniederlassungen berücksichtigen müssen, kann es oft notwendig sein, mehrere Zwischenschritte zu definieren.

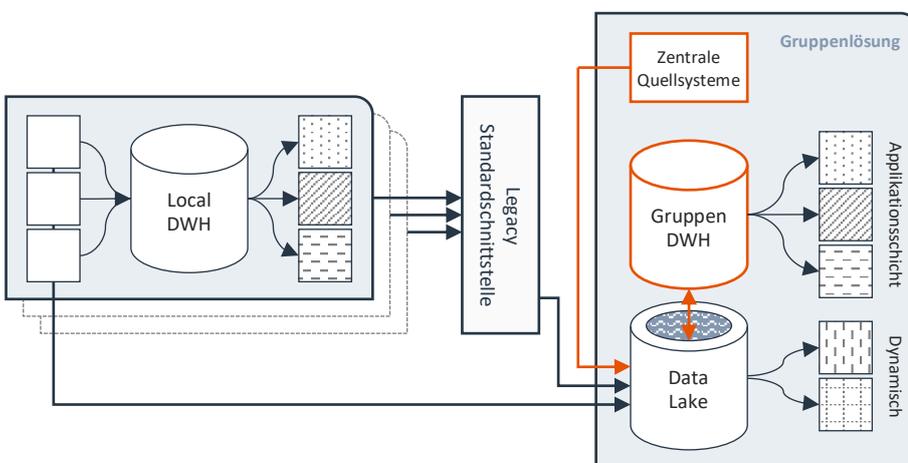
Die folgende Darstellung zeigt diese Idee schematisch vereinfacht am Beispiel eines unserer Kunden, einer international tätigen Bankengruppe:



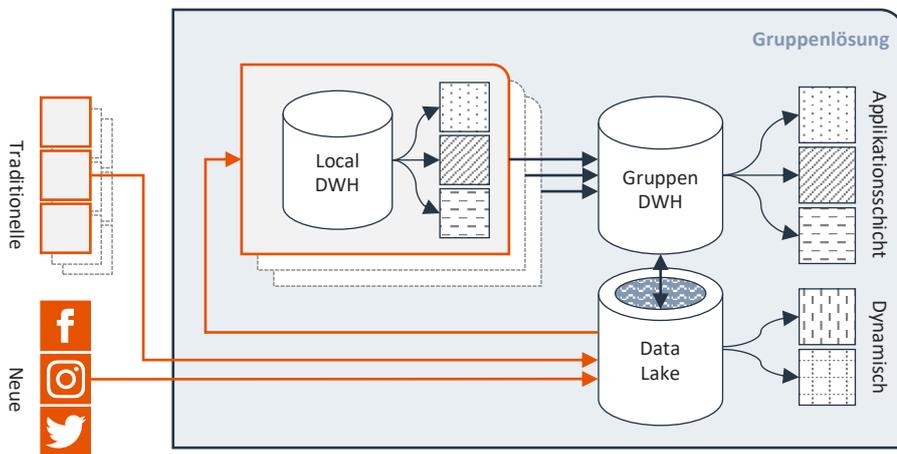
Schritt 1: stellt die Ausgangslage vor Projektbeginn dar, in der die Töchter über eine Standardschnittstelle an das Gruppen-DWH liefern, welches auch die Daten der zentralen Quellsysteme hält.



Schritt 2: führt einen auf Hadoop basierten Data Lake ein, der sowohl aus der Standardschnittstelle als auch den lokalen Quellsystemen beliefert wird. Die Verarbeitung auf Gruppenebene bleibt vorerst davon unberührt, der Data Lake wird aber bereits den Fachbereichen zur Verfügung gestellt.



Schritt 3: der Data Lake wird die neue Akquisitionsschicht der DWH-Architektur, das Gruppen-DWH wird auf die neue Lösung umgestellt. Auch die zentralen Quellsysteme werden ab jetzt über den Data Lake ins DWH gebracht.



Schritt 4: die lokalen DWH werden in die Gruppenlösung integriert, neben den traditionellen Datenquellen werden auch neue Quellen wie bspw. soziale Medien erschlossen.

Wie aus dem Beispiel hoffentlich ersichtlich wird, müssen die jeweiligen Änderungen nicht unbedingt groß oder für alle am Vorhaben Beteiligten relevant sein, um eine Zwischenschritt zu rechtfertigen. Vielmehr geht es darum, in mitunter kleinen, aber schnellen Schritten zusätzliche Daten oder neue Auswertungs- und Berichtsmöglichkeiten zur Verfügung zu stellen und ein jahrelanges Warten auf den „Big Bang“ zu vermeiden.

Ein beliebtes Mittel in dieser Hinsicht ist, auch in Verbindung mit dem nachfolgenden Abschnitt, den Fachbereichen Zugriff auf die oft nur schwer zugänglichen Rohdaten in Form eines Data Lakes für explorative Analysen zur Verfügung zu stellen, während parallel die aufwändigere Datenintegration in Angriff genommen wird.

SCHNELLE ERGEBNISSICHERBARKEIT

Nie hängen die Früchte tatsächlich so tief wie in den bunten Präsentationen, die als Entscheidungsvorlage dienen. Kein Anbieter hat bisher den heiligen Gral entdeckt, der schnell und einfach das möglich macht, was bisher aufwändig und teuer war.

Trotzdem zahlt es sich aus, sorgfältig ein Pilotthema, z.B. eine Applikation, die ohne strukturelle Veränderungen auf die neue Plattform migriert wird, oder einen Datenbereich, für den der „Durchstich“ erfolgt, auszuwählen.

Auch wenn der Nutzen aus der Sicht der IT oft überschaubar ist, so darf man doch nicht die psychologische Wirkung der Sichtbarkeit von Ergebnissen unterschätzen, wenn es darum geht, Entscheidungsträger ins Boot zu holen und auch dort zu halten. Hier ist einem pragmatischen Ansatz ganz klar der Vorzug gegenüber der reinen Lehre zu geben.

MODELLGETRIEBENE ENTWICKLUNG

In den letzten Jahren macht verstärkt das Schlagwort der „DWH-Automatisierung“ die Runde: nicht nur hochgelobte Start-Ups, auch jeder namhafte Hersteller bietet diesbezüglich Funktionalität an. Unseres Erachtens sollten dahingehende Versprechungen sorgfältig geprüft werden, der Projektalltag zeigt zumeist nur begrenzte Zeit- und Aufwandsparnis.

Ein Ansatz, der tatsächlich von zahlreichen Unternehmen erfolgreich angewandt wurde, ist die modellgetriebene Entwicklung, für die das weiter oben bereits erwähnte, fachliche Datenmodell eine hervorragende Ausgangsbasis darstellt.

In Kundenprojekten haben wir nicht nur technische Datenmodelle daraus abgeleitet, sondern auch die Generierung der ETL-Strecken erheblich beschleunigt. Es wäre vermessen, von einer Vollautomatisierung zu sprechen, aber eine Aufwandsreduktion von bis zu 80 Prozent sind – in Abhängigkeit von Projektumfeld, Modellqualität und Werkzeugunterstützung – möglich.

GOVERNANCE

Da die Anforderungserhebung notwendigerweise in enger Zusammenarbeit mit den Fachbereichen erfolgt, sind Modernisierungsvorhaben die ideale Gelegenheit, die Zuständigkeiten und Verantwortlichkeiten für Daten zu überprüfen und ggf. anzupassen oder überhaupt erstmals alltagstaugliche Strukturen und Prozesse aufzusetzen.

Diejenigen unserer Kunden, die sich von unserem Business Information Modelling Ansatz überzeugen ließen, konnten einen weiteren Nutzen daraus ziehen, denn ein fachliches Datenmodell eignet sich wie kein anderes Instrument dazu, fachliche Zuständigkeiten für die verschiedenen Datenbereiche zu definieren. Damit kann sichergestellt werden, dass Data Governance nicht zum „Papiertiger“ verkommt, sondern Nutzen für die Organisation stiftet.

ZUSAMMENFASSUNG

Um die zu Beginn gestellte Frage zu beantworten: nein, das DWH ist nicht tot; ganz im Gegenteil, es erfreut sich vielerorts bester Gesundheit – oder ist zumindest auf dem Weg der Besserung. Die neuen Technologien stellen gewöhnlich eine wertvolle Ergänzung, aber fast nie einen gleichwertigen Ersatz für das DWH klassischer Prägung dar. Noch haben wir z.B. keinen CFO getroffen, der ausschließlich mit Daten aus einem Data Lake den Jahresabschluss erstellen möchte.

Die Modernisierung des DWH wird dann gelingen, wenn sie nicht vorrangig oder gar ausschließlich als IT-Aufgabe gesehen wird, sondern unter enger Einbeziehung der Fachbereiche erfolgt. Ein weiterer Vorteil daraus kann ein Eindämmen des IT-Wildwuchses in den Fachbereichen sein.

Ebenso wichtig ist ein behutsames Vorgehen, denn der große Wurf gelingt nur selten. Viel eher führt ein schrittweiser Ansatz, der auf einem soliden, datengetriebenen Fundament steht zum Erfolg. Wenn es darüber hinaus noch gelingt, eine zeitgemäße Data Governance zu definieren, steht dem DWH noch ein langes, erfülltes Leben bevor!

ÜBER SIMPLITY

Simplify bietet Dienstleistungen rund um Datenmanagement und Analytics und hat zahlreiche führende Finanzinstitute im zentralen europäischen Raum bei der erfolgreichen Umsetzung solcher Initiativen unterstützt.

Das Leistungsspektrum reicht von strategischer Beratung in den Bereichen Architektur, Data Governance und Datenqualitätsmanagement bis zu technischen Implementierungen von Data Warehouses, Data Lakes und Stammdatenmanagement-Lösungen.

Das hauseigene Accurity Softwareprodukt kann dabei sowohl als unterstützendes Werkzeug in Simplify-Projekten als auch als eigenständige Data-Governance-Lösung für Metadaten, Datenqualitäts- und Referenzdatenmanagement zum Einsatz kommen.

AUSGEWÄHLTE KUNDEN



DIE AUTOREN



STEFAN RUHLAND

Principal Consultant bei Simplify und verantwortet die fachliche Seite

zahlreicher Umsetzungsprojekte. Zuvor war er für verschiedene Beratungshäuser und IT-Dienstleister tätig.

stefan.ruhland@simplify.eu

[+436766486918](tel:+436766486918)



STEFAN MARKUS

Senior Consultant bei Simplify und als leitender Datenarchitekt für Kunden

unterschiedlicher Branchen tätig. Seine Laufbahn begann er als Datenmodellierer und BI Consultant bei Teradata.

stefan.markus@simplify.eu

[+436765756950](tel:+436765756950)

KONTAKT



Simplify AT GmbH
Hainburger Straße 20/10
1030 Wien, Österreich
www.simplify.eu | www.accurity.eu

