

Data Warehouse

(DWH)



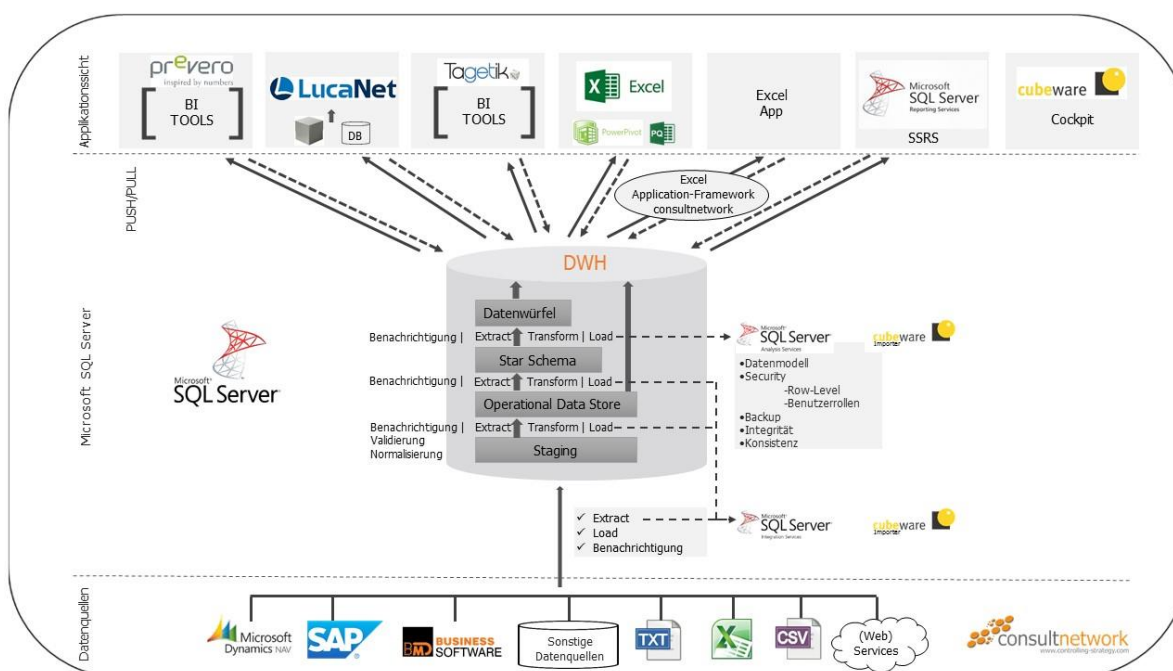
Wien | Twin Tower: Wienerbergstraße 11/12a | 1100 Wien
Klagenfurt am Wörthersee | Bahnhofstraße 49 | 9020 Klagenfurt
Rosenheim | Kufsteiner Straße 103 | 83026 Rosenheim
WWW.CONSUFTNETWORK.COM

Inhaltsverzeichnis

1	Einleitung	2
2	Definition eines Data Warehouse	3
2.1	Aufbau eines Data Warehouse.....	3
2.1.1	Das Staging	4
2.1.2	Der Operational Data Store (ODS).....	4
2.1.3	Das Star Schema.....	4
2.1.4	Der OLAP Datenwürfel	5
2.2	Datenmodellierung-Historisierungskonzept.....	6
2.2.1	Slowly Changing Dimensions	6
2.3	Normalisierung.....	7
3	Organisationsformen eines Data Warehouse	8
3.1	Zentrales Data Warehouse.....	8
3.2	Dezentrales Data Warehouse / Data Marts.....	8
3.3	Virtuelles Data Warehouse	8
3.4	Hub-and-Spoke Architektur	9
4	Datenverteilung im Data Warehouse	10
4.1	Das Pull - Prinzip	10
4.2	Das Push - Prinzip	10
5	Analyse- und Präsentations-Schicht.....	11
5.1	Sandbox.....	12
6	Softwareunterstützung für den Aufbau eines Data Warehouse	13
6.1	Microsoft Programmlösungen: SQL-Servers.....	13
6.1.1	SQL Server Integration Services (SSIS)	13
6.1.2	SQL Server Analysis Services (SSAS).....	13
6.1.3	SQL Server Reporting Services (SSRS)	13
6.2	Cubeware Programmlösung: Cubeware Solutions Platform	14
6.2.1	C8 Importer	14
6.2.2	C8 Server.....	14
6.2.3	C8 Cockpit	15
7	Implementierungsvarianten Data Warehouse	16
7.1	Umfassendes DWH	16
7.2	Bedarfsorientiertes DWH.....	16
8	Zusammenfassung	17
9	Die Autoren.....	18
	Mario Rosenfelder	18
	Bernhard Rosenfelder	18

1 Einleitung

In Ihrem Unternehmen liegen Unmengen an heterogenen Daten, aus unterschiedlichsten Systemen mit jeweils unterschiedlichen Datenständen und ohne logische Verknüpfungen vor? Dann sind Sie möglicherweise an dem Punkt angekommen, an dem die Einführung eines zentralen Data Warehouse für die Zukunft Ihres Unternehmens erfolgskritisch ist! Ein Data Warehouse ist eine zentrale Sammelstelle für internen und externen Daten aus unterschiedlichen Datenquellen. Im Datawarehouse werden die Daten aus verschiedenen Systemen logisch zusammengeführt und Inkonsistenzen beseitigt. Darüber hinaus kann eine Automatisierung der Datengewinnung und Bereitstellung realisiert werden. Das Data Warehouse liefert einen unternehmensweit einheitlichen Datenbestand, auf den alle Mitarbeiter entsprechend ihren Rechten zugreifen können.



Wie aus der Graphik ersichtlich, werden die Daten aus verschiedensten Vorsystemen extrahiert (E) und in die unterste Datenebene das Staging des DWHs geladen (L). Das Staging ist ein mehr oder weniger reduzierter Datenbestand aus den Vorsystemen der in weiterer Folge noch mal extrahiert, transformiert (T) und in den sogenannten Operational Data Store geladen wird. Man spricht in diesem Zusammenhang von einem sogenannten ETL-Prozess. Bei großen Datenmengen macht es Sinn, aufbauend auf dem Operational Data Store noch ein Star Schema welches die Basis für die Erstellung multidimensionaler Datenwürfel bildet zu erstellen. Die Datenwürfel haben den Vorteil, dass Sie eine sehr schnelle Auswertung großer Datenmengen ermöglichen. Die Daten werden daraufhin entweder über die Datenwürfel oder aus dem Operational Data Store heraus verteilt und damit für jeden zugänglich gemacht. Somit kann gewährleistet werden, dass die im Warehouse befindlichen Daten mit verschiedenen Applikationen ausgewertet werden können. Ziel eines zentralen Data Warehouses ist es, durch die Verfügbarkeit von internen und externen Informationen Vorteile gegenüber dem Wettbewerb zu erzielen!

consultnetwork hilft Ihnen bei der Konzeptionierung Ihres unternehmensweiten Data Warehouses und stellt Ihnen vorab mit diesem kostenlosen Whitepaper einige wissenswerte Informationen zum Thema Data Warehouse zur Verfügung. Sollten Sie Fragen haben wenden Sie sich an:

Kontakt:

Ing. Mag. Mario Rosenfelder (CEO)

+49 (0)8031 58180-12

+43 (0)463 219095-12

mario.rosenfelder@consultnetwork.com

2 Definition eines Data Warehouse

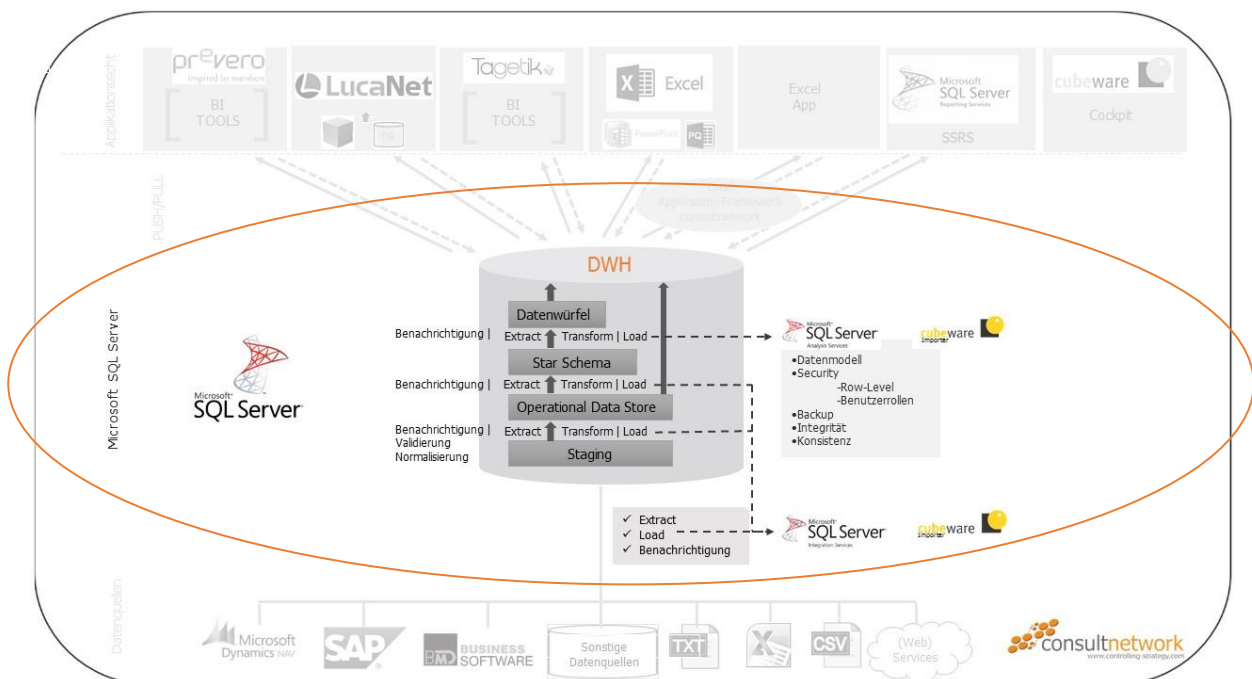
Ein Data Warehouse ist eine zentrale Sammelstelle für interne und externe Daten. Es integriert Informationen aus verschiedenen Datenquellen in eine für die Entscheidungsfindung optimierte Datenbank. Die Daten werden aus unterschiedlichsten Vorseystemen extrahiert, wenn erforderlich transformiert und danach in das Data Warehouse geladen und dort dauerhaft gespeichert.

Man spricht in diesem Zusammenhang von einem sogenannten ETL-Prozess, also Extract-Transform-Load-Prozess, der beim Aufbau der verschiedenen Datenebenen mehrfach durchlaufen wird:

- Extraktion relevanter Daten aus Quellsystemen/Datenebenen,
- Transformation und Bereinigung der Daten aus Vorseystemen/Datenebenen,
- Laden der bereinigten, konsistenten Daten in das DWH (relational=ODS, multidimensional=Datenwürfel)

2.1 Aufbau eines Data Warehouse

Folgende Abbildung zeigt den Aufbau eines Data Warehouse. In der Regel realisieren wir ein Warehouse mit Hilfe von vier verschiedenen Datenebenen. Die unterste Ebene ist das Staging, danach folgt der Operationale Data Store, kurz ODS. Auf den ODS aufbauend folgt das Star Schema. Im Star Schema werden die Daten so bereitgestellt, dass der Aufbau der obersten Ebene: der Datenwürfel, erfolgen kann. Die einzelnen Ebenen werden nachfolgend genauer erklärt, um Ihnen einen Einblick in die Konzeption eines Data Warehouse zu gewährleisten.



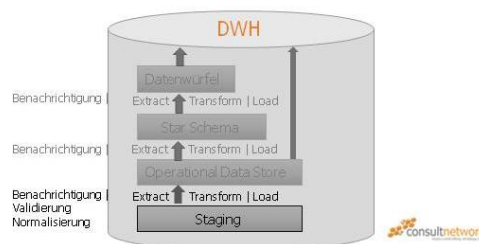
2.1.1 Das Staging

Das Staging ist die unterste Ebene des Data Warehouse. In das Staging werden alle relevanten Daten aus den Vorsystemen des Unternehmens und aus den externen Datenquellen hinein extrahiert. Das Staging hält im Wesentlichen ein Abbild der extrahierten Daten aus den Vorsystemen.

Die Daten werden je nach Bedarf täglich, stündlich oder sogar jede Minute geladen. In weiterer Folge werden dann alle anderen Datenebenen und Datenempfänger aus dem Staging heraus bedient. Das hat den Vorteil, dass nicht mehr jede Planungs-, Konsolidierungs- oder BI-Applikation selbstständig auf die Vorsysteme zugreifen muss. So wird verhindert, dass sich die Vorsysteme aufgrund unnötiger Mehrfachzugriffe verlangsamen, bzw. garantiert, dass ein einheitlicher Datenbestand vorherrscht.

Mit verschiedenen Tools kann die Extraktion und das Laden der Daten in das Data Warehouse automatisiert werden. Solche Werkzeuge sind beispielsweise das SQL Server Integration Service oder der Cubeware Importer.

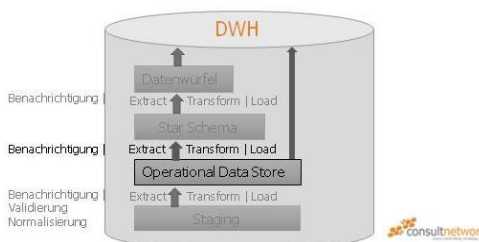
Umfassendere Informationen zu den Themen Cubeware und SQL Server Integration Service finden Sie unter <http://www.controlling-strategy.com>. Wenn Sie noch weitere Informationen zu den Themen Microsoft BI oder Cubeware benötigen, dann fordern Sie doch gleich jetzt unser kostenloses [Whitepaper Microsoft BI](#) bzw. Whitepaper Cubeware an.



2.1.2 Der Operational Data Store (ODS)

Aus der Datenebene Staging werden Daten im nachfolgenden Schritt in den Operational Data Store, kurz ODS, extrahiert, transformiert und geladen. Im Zuge dieses Vorgangs werden Redundanzen und Integritätsverletzungen beseitigt und die Datenbestände unterschiedlicher Vorsysteme miteinander verknüpft.

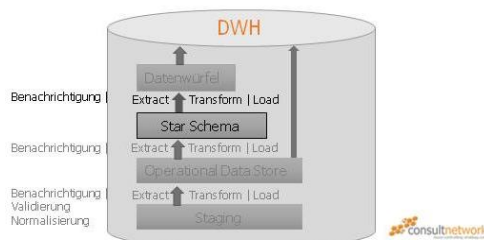
Auf der Ebene des Operational Data Store können bereits Daten für aktuelle und zeitnahe Auswertung bereitgestellt werden. Verschiedenste Software Tools wie zum Beispiel Tagetik oder das von uns entwickelte [Excel Application Framework](#) können bereits direkt auf die gespeicherten Daten im ODS zugreifen. Die Datenhaltung im Operational Data Store (ODS) erfolgt mit Hilfe eines relationalen Datenmodells. Vor allem bei großen Datenmengen können Abfragen auf den Operational Data Store langsam werden und zu langen Antwortzeiten führen. Hier empfiehlt es sich multidimensionale Datenwürfel einzusetzen.

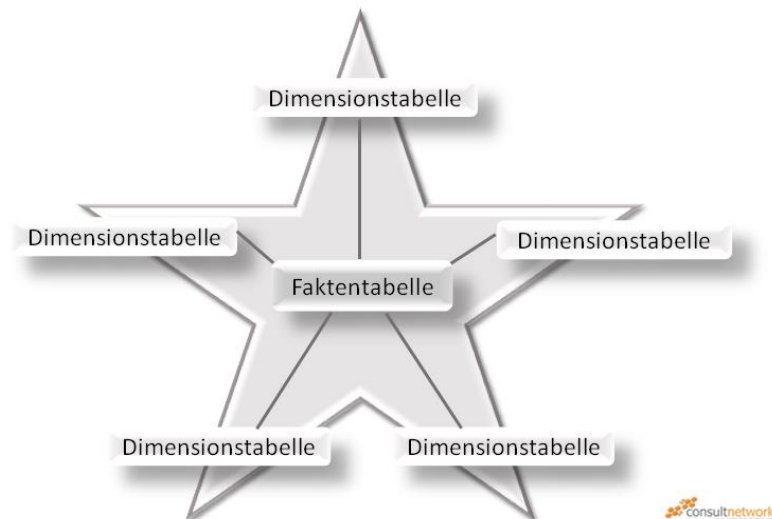


2.1.3 Das Star Schema

Die dritte Datenebene in unserem DWH Modell ist das Star Schema. Diese Ebene wird in vielen Fällen ergänzend zum Operational Data Store aufgebaut und unterstützt in weiterer Folge den Aufbau der Datenwürfel, welche vor allem bei großen Datenmengen benötigt werden.

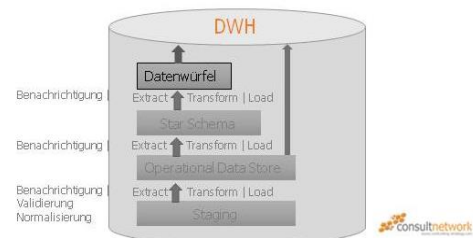
Im Star Schema werden die Daten so bereitgestellt, dass der Aufbau des Datenwürfels mit seinen Dimensionen optimal unterstützt wird. Das Star Schema hat zwei Tabellentypen: die Dimensionstabelle und die Faktentabelle (Bewegungsdaten). Im Zentrum befindet sich eine Faktentabelle mit Kennzahlen, Metriken und Messwerten. Um die Faktentabelle näher zu beschreiben, werden Tabellen sternförmig um sie herum angeordnet. Diese werden als Dimensionstabellen bezeichnet. Die einzelnen Informationen der Dimensionstabelle definieren in ihrer Gesamtheit wiederum die Faktentabelle. Auf Grund dieser Anordnung wird diese Datenebene Star Schema genannt.





2.1.4 Der OLAP Datenwürfel

Der Datenwürfel ist die oberste Datenebene unseres Data Warehouse-Modells. Die Daten sind dabei als Elemente eines mehrdimensionalen Würfels zu verstehen. Die Dimensionen - aufgebaut aus den Tabellen des Star Schemas - beschreiben die Daten und ermöglichen einen einfachen Zugriff auf die Daten, indem die Datenpunkte über die einzelnen Elemente der Achsen des Würfels gewählt werden. Der Datenwürfel wird auch als OLAP-Würfel bezeichnet, wobei OLAP für Online Analytical Processing steht.



Der OLAP-Würfel ist eine mehrdimensionale Darstellung von Kennzahlen. Der Datenwürfel besteht aus verschiedenen Datenzellen, die eine oder mehrere Kennzahlen beinhalten. Er wird durch Kombination der Dimensionen konstruiert und besteht somit aus einzelnen Datenzellen. Diese Zellen stellen die Schnittpunkte der Dimensionen (z.B. Monate, Vertriebsregionen, Kunden, Produkte, ...) dar und beinhalten eine oder mehrere Kenngrößen (z.B. Umsatz, Gewinn, Deckungsbeitrag, Rendite, Gewicht, Länge, ...).

Um den Datenwürfel analysieren zu können, werden verschiedene Applikationen, wie Excel, das Cube-ware Cockpit oder die Reporting Services des SQL-Server (SSRS), herangezogen. Im Zuge des Data Minings werden die Daten für die Analyse aufbereitet und gefiltert. Somit sind die Daten für die Analyse optimal vorbereitet. Die oben genannten Applikationen können aber auch Daten an das Data Warehouse zurückschreiben. Dadurch sind auch die Plandaten im zentralen Warehouse unternehmensweit verfügbar.

Der gesamte Verlauf der Datenbeschaffung, Verwaltung und Auswertung eines Data-Warehouses wird auch als **Data Warehousing** bezeichnet. Hierzu gehören folgende Aufgaben:

- die Datenbeschaffung, sowie die Integration der Daten und deren Weiterverarbeitung,
- die Datenhaltung, sprich die langfristige Speicherung der Daten,
- die Datenauswertung, sowie die Analyse der Daten,
- die Versorgung und Datenhaltung der für die Analyse notwendigen separaten Datenbestände, den Data-Marts.

Den gesamten Prozess des Data Warehouse Aufbaus realisieren wir mit dem Cubeware Importer oder den On-Bord-Werkzeugen des SQL-Servers (SSIS und SSAS). Wollen auch Sie ein Data Warehouse in Ihrem Unternehmen implementieren oder haben Sie Fragen zu diesem Thema, dann wenden Sie sich bitte an:

Kontakt:

Ing. Mag. Mario Rosenfelder (CEO)

+49 (0)8031 58180-12

+43 (0)463 219095-12

mario.rosenfelder@consultnetwork.com

2.2 Datenmodellierung-Historisierungskonzept

Unternehmen müssen die Handhabung mit einer Vielzahl an Daten beherrschen. Alle unternehmensweiten Daten gehören erhoben, verwaltet und interpretiert, was eine große Herausforderung darstellt. Ein Data-Warehouse unterstützt Ihr Unternehmen dabei!

Informationen aus verschiedenen Quellen werden in das Staging integriert, damit die Daten mit Hilfe des ETL-Prozesses weiterverarbeitet und daraufhin analysiert werden können. Auch historische Daten spielen hierbei eine wesentliche Rolle.

Wie wir bereits wissen, kann bei einem Data-Warehouse zwischen Faktentabelle und Dimensionstabelle unterschieden werden. Die Faktentabelle beinhaltet Bewegungsdaten aus Quelldatenbanken. Diese Daten verändern sich sehr oft, wodurch hier häufig Aktualisierungen stattfinden. Historische Daten befinden sich meistens in einer Dimensionstabelle, da sich die in der Dimensionstabelle beschriebenen Daten im Laufe der Zeit kaum verändern. (Bsp.: Information zu einem Produkt) Natürlich kann auch der Fall aufkommen, dass eine Änderung eintritt. (Bsp.: Umbenennung eines Produktes) Um den seltenen Änderungen innerhalb einer Dimensionstabelle gerecht zu werden und diese sinnvoll zu verwalten, gibt es das Prinzip des Slowly Changing Dimensions.

2.2.1 Slowly Changing Dimensions

Slowly Changing Dimensions, kurz SCD, bedeutet ins Deutsche übersetzt „sich langsam verändernde Dimensionen“. Slowly Changing Dimensions sind Methoden, welche Änderungen in Dimensionstabellen erfassen und gegebenenfalls historisch dokumentieren. Hierbei kann zwischen drei Verfahren unterschieden werden, die in Typen unterteilt sind. Was alle drei Typen gemeinsam haben ist, dass vorhandene Datensätze über den Primärschlüssel mit neuen Datensätzen verbunden werden und damit Änderungen in der Tabelle gespeichert werden können.

Die drei verschiedenen Ansätze, die je nach Anforderung an ein Data-Warehouse ausgewählt werden können, werden nun kurz erläutert:

Typ 1: Überschreiben der Werte

Der Typ 1 ist der Typ mit der einfachsten Vorgehensweise, da der neue Wert einfach im Datensatz überschrieben wird. Das bedeutet, dass durch das Überschreiben der ursprüngliche Inhalt verloren geht. Dementsprechend findet hier keine Historisierung statt.

Dieser Typ sollte in Betracht gezogen werden, wenn Informationen einfach nur berichtet werden sollen, ohne die Ursprünglichen zu historisieren.

Typ 2: Hinzufügen eines neuen Datensatzes

Bei diesem Typ werden die vorgenommenen Änderungen historisiert. Hinzugefügt werden bei diesem Ansatz zwei weitere Attribute, die die Gültigkeit eines Datensatzes abbilden. Mit Hilfe eines Datums wird festgehalten, bis wann ein Datensatz gültig bzw. aktuell ist. Wenn notwendig, kann durch ein drittes Attribut diese Aktualität noch zusätzlich gekennzeichnet werden, bspw. mit einer Markierung „wahr“ oder „falsch“. Bei Typ 2 wird also eine Zeile mit der neuen Information hinzugefügt, dadurch geht die jeweilige Veränderung nicht verloren, sondern bleibt in der Dimension bestehen.

Dieser Typ ist der populärste und sollte in Betracht gezogen werden, wenn Informationen über Änderungen hinweg gespeichert, also historisiert werden sollen.

Typ 3: Hinzufügen einer neuen Spalte

Die zweite Art und Weise wie Daten historisiert werden können, stellt Typ 3 des SCD Prinzips dar. Anders als beim Typ 2 werden hier Attribute mit den neuen Werten überschrieben. Die alte Information wird in einer zusätzlichen Spalte in einem Datensatz gespeichert. Nachteil bei dieser Art der Historisierung ist, dass jeweils nur der neue und der alte Zustand gespeichert wird. Das heißt also, dass nur die unmittelbar aktuelle Änderung historisiert wird. Bei häufigeren Änderungen werden die Werte in den entsprechenden Spalten überschrieben und die vorherige Historisierung geht verloren.

Dieser Typ sollte in Betracht gezogen werden, wenn Informationen innerhalb einer Dimension nur sehr selten geändert werden.

Anzumerken ist, dass sich in der Praxis Informationen in einer Dimension häufiger verändern, als in dem Prinzip des Slowly Changing Dimensions dargestellt (Bsp.: Preisänderungen). Gerade bei der Vorgehensweise des Typs 3 kann es dadurch zu einem Nachteil kommen, da durch häufige Änderungen immer die jeweils vorherige Historisierung unsichtbar wird. Typ 3 ist dann anzuwenden, wenn nur selten Änderungen eintreten und ein Vergleich der unmittelbaren vorherigen Veränderung in einem Bericht erwünscht ist.

2.3 Normalisierung

Wird von Normalisierung eines relationalen Datenschemas (Tabellenstruktur) gesprochen, verstehen wir darunter eine Aufteilung von Tabellenspalten in mehrere Tabellen gemäß den Normalisierungsregeln. Dadurch entsteht eine Form, die keine vermeidbaren Redundanzen mehr enthält. Redundanzen sind Mehrfachspeicherungen gleicher Information, die es bei der Normalisierung zu verhindern gilt, um ein Auftreten von inkonsistenten Daten (Anomalien) innerhalb einer Datenbank zu vermeiden. Abhängig von der Art der Anomalie in einem Datenbankschema spricht man von der ersten, zweiten oder dritten Normalform.

Von der **ersten Normalform** wird gesprochen, wenn die gesamte Information in einer Tabelle atomar dargestellt wird.

Eine Tabelle befindet sich in der **zweiten Normalform**, wenn diese in der ersten Normalform ist und alle Nichtschlüsselattribute von allen Schlüsselkandidaten voll funktional abhängig sind. Die 2NF reduziert Redundanzen und die Gefahr von Inkonsistenzen, da jede Relation nur einen Sachverhalt abbildet. Das Verständnis der Datenstruktur wird somit erleichtert und es sind nur noch logisch zusammengehörige Informationen in einer Relation aufzufinden

Eine Datenbank befindet sich in der **dritten Normalform**, wenn die zweite Normalform erfüllt ist und keine transitiven Abhängigkeiten bestehen. Transitive Abhängigkeiten bestehen dann, wenn ein Nichtschlüsselattribut transitiv vom Kandidatenschlüssel abhängig ist. Daraus folgt, dass sich eine Tabelle in der zweiten Normalform befinden muss und kein Nichtschlüsselattribut transitiv von einem Kandidatenschlüssel abhängen darf, um in der dritten Normalform zu sein.

In unseren Data-Warehouse Projekten fokussieren wir uns vor allem auf die dritte Normalform. Denn durch die 3NF werden transitive Abhängigkeiten entfernt und verbliebene thematische Durchmischungen in einer Relation beseitigt, sodass nach der 3NF alle Relationen des Schemas monothematisch sind. Die dritte Normalform ist demnach das Ziel einer erfolgreichen Normalisierung in einem relationalen Datenbankmodell. Diese Normalform verhindert einerseits Anomalien und Redundanzen in Datensätzen und bietet andererseits trotzdem genügend Performance für SQL-Abfragen. So erlangen Sie eine optimierte Datenbankstruktur mit möglichst konsistenten Datenbeständen und möglichst einfacher Datenstruktur.

3 Organisationsformen eines Data Warehouse

Es gibt verschiedene Formen, wie ein Data Warehouse organisiert werden kann. Je nach Anforderungen des jeweiligen Unternehmens kann das Warehouse als zentrale oder dezentrale Datenbasis für managementunterstützende Systeme implementiert werden. In Ausnahmefällen kann auch ein virtuelles Data Warehouse Anwendung finden.

3.1 Zentrales Data Warehouse

In einem zentralem Data Warehouse werden alle Daten an einem Ort gespeichert. Dort werden sie von einem einzigen Datenbank-Management-System kontrolliert und verwaltet. Das bedeutet jedoch nicht, dass die Daten physisch an einem bestimmten Ort gelagert werden müssen. Sie können auch verteilt vorgehalten werden. In diesem Fall muss aber ein verteiltes Datenhaltungssystem die zentrale Verwaltung der verteilten Daten übernehmen.

Ein zentrales Data Warehouse kann jedoch auch einige Nachteile mit sich ziehen. Es können beispielsweise Skalierbarkeitsprobleme aufgrund der zunehmenden Benutzerzahlen und Datenvolumina entstehen und die Performance des Gesamtsystems könnte beeinträchtigt werden. Eine grundlegende Konzeption eines zentralen Data Warehouse ist demnach komplex und benötigt die Hilfe von Experten.

3.2 Dezentrales Data Warehouse / Data Marts

Viele Unternehmen bevorzugen eine dezentrale Ausrichtung ihres Data Warehouses. Ein solch dezentrales Warehouse besteht aus mehreren isolierten, kleineren Data Warehouses, den sogenannten Data Marts. Unter einem Data Mart wird die Kopie eines Teildatenbestands des Data Warehouses verstanden, welcher für einen bestimmten Organisationsbereich oder für eine bestimmte Anwendung bzw. Analyse erstellt wird. Jedes Data Mart besitzt eine eigene Datenhaltung. Demzufolge wird ein Data Mart auch häufig als lokales Data Warehouse bezeichnet, da die Charakteristiken einer Data Mart-Datenbank die gleichen sind, wie bei einem Datawarehouse. Ein Data Mart kann auch als Data View bezeichnet werden.

Sie werden oft zu Beginn der Entwicklung eines zentralen, unternehmensweiten Data Warehouse verwendet. Ein Data Mart ist also ein physikalisch getrennter Auszug aus einem Data Warehouse, dessen Datenbestand einer Abteilung für spezifische Auswertungen zur Verfügung gestellt wird. Diese spezifische Auswertung wird mit Hilfe von OLAP- oder Data Mining-Werkzeugen durchgeführt. Hauptgrund für die Erstellung von Data Marts ist die Tatsache, dass sie die Möglichkeit bieten die Verarbeitungsgeschwindigkeiten zu erhöhen. Darüber hinaus kann die Datenstruktur an die Anforderungen der jeweiligen Abteilung angepasst werden.

Aber auch ein dezentrales Data Warehouse hat Nachteile. Denn mit einer dezentralen Ausrichtung geht auch die integrierte Sichtweise auf das Gesamtunternehmen verloren. Das Durchführen von unternehmensweiten Analysen wird erschwert.

3.3 Virtuelles Data Warehouse

Beim rein virtuellen Data Warehouse wird nicht auf eine spezielle Datenbank zugegriffen. Der Zugriff erfolgt direkt über die operativen Systeme, was zu einer starken Belastung der Performance und zu einer schlechten Datenqualität führt. Aufgrund dessen wird diese Organisationsform nur in Ausnahmefällen oder als Ergänzung zu einem klassischen Data Warehouse verwendet, nämlich dann, wenn ein direkter Zugriff auf eine begrenzte Menge von operativen, hochaktuellen Daten zur Verfügung gestellt werden soll.

Interessieren Sie sich für die Einführung eines Data Warehouses in Ihrem Unternehmen? Wir beraten Sie gerne weiter und beantworten Ihre Fragen! Kontaktieren Sie uns am besten gleich jetzt:

Kontakt:

Ing. Mag. Mario Rosenfelder (CEO)

+49 (0)8031 58180-12

+43 (0)463 219095-12

mario.rosenfelder@consultnetwork.com

3.4 Hub-and-Spoke Architektur

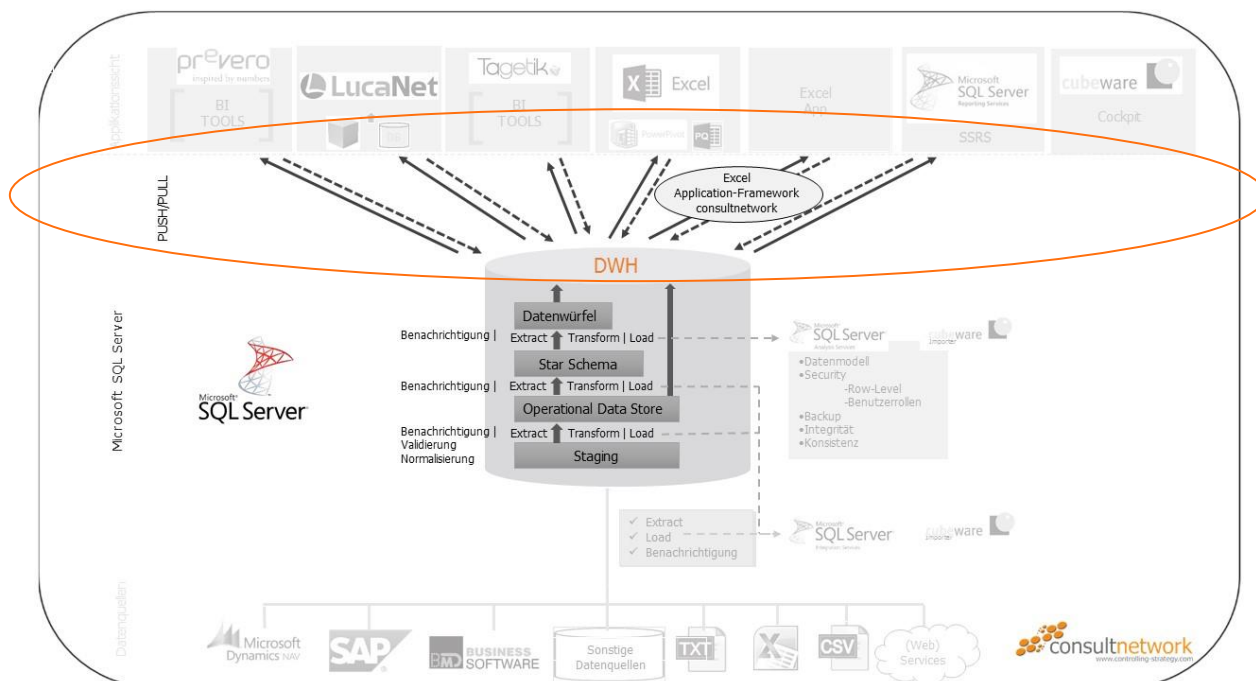
Als Hub-and-Spoke Architektur wird eine Architektur mit zentralem Data-Warehouse und angeschlossenen Data Marts (Kapitel 3.2) charakterisiert. Diese Architektur wird als Hub-and-Spoke bezeichnet, da die Anordnung der Komponenten an eine Naben-Speichen-Kombination erinnert. Bei dieser Architektur werden die zur Verfügung stehenden Datenquellen im Data-Warehouse aus einzelnen Datenquellen zusammengeführt, bereinigt und historisiert. Aus dem konsistenten Datenbestand, dem Hub, werden dann meistens abteilungsbezogene Data Marts, Spokes, erzeugt. Das Data-Warehouse füttert sozusagen die Data Marts mit Daten und Informationen, welche mittels ETL-Prozess dorthin gelangen.

Vorteile dieser Architektur sind die geringere Anzahl an Schnittstellen und eine zentrale Integration und Aufbereitung der Daten. Ein Hauptaugenmerk liegt auf der Transformation in einheitliche (consistent), bereinigte (clean), vollständige (comprehensive), aktuelle (current) und übereinstimmende (conformed) Daten.

Diese Form der Architektur, welche alle Daten auf mehrere Data Marts verteilt und speichert, hat sich als Best-Practice-Anwendung zur Standardisierung eines Data-Warehouses erwiesen.

4 Datenverteilung im Data Warehouse

Daten können im Data Warehouse auf zwei verschiedene Arten verteilt werden. Entweder nach dem Pull-Prinzip oder dem Push-Prinzip.



4.1 Das Pull - Prinzip

Bei dem Pull-Prinzip ist der Datenempfänger, der eine Abfrage durchführt, verantwortlich für die Auswahl der Informationen, sowie den Datentransport. Um zu vermeiden, dass unbefugte Personen Zugriff zu Daten haben, welche nicht für sie bestimmt sind, werden im Vorhinein Berechtigungen vergeben. Auf Basis dieser Berechtigungen wird geregelt, welche Daten aus dem Data-Model von wem abgegriffen werden dürfen. Der Datenempfänger kann lediglich bestimmen wie oft, zu welchem Zeitpunkt und in welcher Struktur er eine Datenabfrage durchführt.

4.2 Das Push - Prinzip

Bei dem Push-Prinzip entscheidet der Lieferant, also das Data Warehouse, welche Daten in welcher Form und zu welchem Zeitpunkt angeboten werden. Dem Datenempfänger bleibt lediglich die Entscheidung über, welche Daten er benötigt. Der Vorteil des Push-Prinzips gegenüber dem Pull-Prinzip liegt darin, dass ein willkürlicher Zugriff auf sensible Daten nicht möglich ist. Der Zugriff wird Unbefugten durch Regelungen verboten. Dieses Prinzip der Datenverteilung macht bei einem großen Data Warehouse besonders viel Sinn. Auch auf Grund dessen, dass die Fehlerquoten und deren Auswirkungen beim Push-Prinzip viel geringer sind als beim Pull-Prinzip.

Im Warehouse können Tabellen, Spalten aber auch eine Row Level Security mit dem SQL Server umgesetzt werden. Auf Basis von bestimmten User-Rollen können eigene SQL-User, als auch Domain-User, spezielle Berechtigungen erhalten. Somit können einzelne Zugriffe geregelt werden. Lese- sowie Schreibzugriffe werden bereits auf der untersten Datenebene des Data Warehouses, dem Staging, geregelt. Dadurch kann ein unberechtigter Zugriff auf die Daten verhindert werden.

Das C8 Cockpit beherrscht diese Web-Technologie wie kein anderes in diesem Segment. Wenn auch Sie sich für solch ein leistungsstarkes C8 Cockpit interessieren, dann besuchen Sie uns auf www.consulting-strategy.com oder wenden sich gleich direkt an uns! Wir liefern Ihnen wissenswerte Informationen zu diesem Produkt.

Kontakt:

Ing. Mag. Mario Rosenfelder (CEO)

+49 (0)8031 58180-12

+43 (0)463 219095-12

mario.rosenfelder@consultnetwork.com

5.1 Sandbox

Das Ziel einer analytischen Sandbox ist es, Business Analytikern und Power Usern die Möglichkeit zu bieten, auf verschiedene Daten aus internen und externen Quellen des Unternehmens zuzugreifen. Mit ihrer Hilfe sollen geeignete Speicher, Werkzeuge und Verarbeitungsressourcen bereitgestellt werden, um den Bedarf an data shadow systems zu verringern. So wird das Durchspielen verschiedenartiger Szenarien, mittels Anreicherung des bestehenden Datensatzes um eigene Daten, ermöglicht.

Einstellungen und Änderungen in der Sandbox sind nur für den jeweiligen User sichtbar. Es besteht aber auch die Möglichkeit, die Berichte basierend auf den angereicherten Daten der Sandbox zu erstellen und anderen Usern zur Verfügung zu stellen.

6 Softwareunterstützung für den Aufbau eines Data Warehouse

6.1 Microsoft Programmlösungen: SQL-Servers

Wie oben bereits beschrieben kann der gesamte Prozess des Data Warehousing mit den On-Bord-Werkzeugen des SQL-Servers realisiert werden. Hierbei unterscheiden wir zwei verschiedene Tools, mit welchen die Implementierung des Warehouses vonstattengehen kann. Dem SSIS, SQL Server Integration Services, und dem SSAS, dem SQL Server Analysis Services.

Die gesamten SQL Server Data Tools (SSDT) bilden im Kontext eine vollständige BI Lösung. Es gibt drei Bereiche (SSIS, SSAS und SSRS), welche jeweils einen Teil eines BI Projektes übernehmen. Nachfolgend werden die einzelnen drei Programme erläutert, welche in ihrer Gesamtheit die SSDT ergeben.

6.1.1 SQL Server Integration Services (SSIS)

Die SQL Server Integration Services bieten eine vollständige Lösung im Bereich der Datenintegration und Datentransformation an. Daten können aus jeder beliebigen Quelle abgeholt, zusammengeführt und verändert werden. Das Programm bietet dafür eine sehr übersichtliche grafische Oberfläche, welche es ermöglicht, einen ganzen Import zu erstellen ohne eine Zeile Code zu verfassen. Natürlich kann bei Bedarf auch auf Scripts zurückgegriffen werden und es können auch eigene Abfragen definiert werden. Die fertigen Pakete können direkt aus Visual Studio heraus auf dem SQL Server bereitgestellt werden. Die bereitgestellten Pakete werden dabei in einem eigenen SSIS Katalog geladen und können über den SQL Server Agent, sowie über die Windows Aufgabenplanung zielgesteuert angestoßen werden. Ein Paket beinhaltet alle Schritte eines Datenimports und ermöglicht eine Ausführung auch außerhalb der Entwicklungsumgebung.

6.1.2 SQL Server Analysis Services (SSAS)

Die SQL Server Analysis Services stellen OLAP und Data Mining Funktionalitäten bereit. Mit dem SSAS kann ein multidimensionales Modell, basierend auf z.B. relationalen Daten erstellt werden. Mit Hilfe von SQL Server Analysis Services lassen sich OLAP Würfeln designen, KPI's können definiert und als neue Measures hinzugefügt werden. Abhängig von der Version des SQL Servers können auch unterschiedliche Sichten des Würfels bereitgestellt werden und individuelle Benutzerrechte vergeben werden.

Nachteil: Bei der Anwendung gilt es jedoch zu bedenken, dass die daraus erstellten OLAP-Würfel für die Analysis Services konzipiert sind und nicht auf anderen Datenbanken bereitgestellt werden können. Sollten Sie eine andere OLAP Datenbank verwenden, können Sie sich bei uns unter www.controlling-strategy.com über die Alternativen informieren.

6.1.3 SQL Server Reporting Services (SSRS)

Mit dem SQL Server Reporting Services können Berichte und Dashboards generiert und über ein Webinterface bereitgestellt werden. Die Berichte können auch interaktiv gestaltet werden, sodass Benutzer individuelle Einstellungen vornehmen können. Ein Rollensystem garantiert für die Sicherheit der Daten am Reporting Server - natürlich können bestimmte Berechtigungen auch vererbt werden. Statische Reports können durch eine Abonnement-Funktion in regelmäßigen Abständen erzeugt werden. Das Besondere an den SQL Server Reporting Services (SSRS) ist die Möglichkeit einer datengesteuerten Abonnement Funktion, die es erlaubt, Berichte automatisch zu erstellen und per E-Mail zu verschicken. Alternativ können die Berichte auch in Share Point geladen werden - dabei wird die Verteilung und Darstellung durch Share Point vorgenommen. Der SSRS kann, wie im vorhergehenden Kapitel bereits angesprochen zu der obersten Architekturschicht eines Datawarehouse gezählt werden und dient der Analyse und Präsentation von Daten.

Nachteil: Die Berichte der Reporting Services sind in erster Linie auf die Visualisierung ausgelegt und nicht dafür gedacht, Daten zurückzuschreiben. Theoretisch könnte man dies zwar ermöglichen, jedoch

nur in beschränkter und nicht benutzerfreundlicher Form. Aus diesem Grund sind sie für Planungszwecke eher ungeeignet. Außerdem sind die grafischen Darstellungsmöglichkeiten etwas überholt.

Die **SQL Server Data Tools** bieten eine vollständige Lösung um alle Aufgaben eines BI Projektes zu bewältigen und das in einer integrierten Umgebung. Auch komplexe Anforderungen lassen sich damit leicht bewältigen. Die gewonnenen Daten können zu einem umfassenden Data Warehouse zusammengefasst werden.

Der Vorteil gegenüber anderen Produkten ist das nahtlose Zusammenspiel aller Tools in einer einzigen Umgebung. Sie können ein Paket für den Datenimport erstellen, einen OLAP Würfel erstellen, abschließend die dazugehörigen Berichte designen, über einen Server zur Verfügung stellen und das ohne ein einziges Mal die Entwicklungsumgebung wechseln zu müssen. Dieses Tool deckt alle Teilprozesse eines umfassenden BI Projektes ab und ist bereits automatisch in SQL Server enthalten.

Da in sehr vielen Unternehmen sowohl Excel, SQL-Server und auch Sharepoint bereits vorhanden sind, ist es immer wieder möglich, BI Projekte mit einem geringen Investment in Lizenzen zu realisieren. Sollte einer der angeführten Nachteile für Ihren Anwendungsfall erfolgskritisch sein, kontaktieren Sie uns. Wir haben Lösungen und Tools im Angebot, die genau darauf abzielen diese Schwächen zu beheben! Des Weiteren steht Ihnen auch unser kostenloses [Whitepaper Microsoft BI](#) zur Verfügung, welches sie gleich kostenlos anfordern können.

6.2 Cubeware Programmlösung: Cubeware Solutions Plattform

Die Cubeware Solutions Plattform C8 hilft Ihnen dabei alle notwendigen Daten und Informationen, die sich in Ihrem Unternehmen befinden, zusammenzufügen. Diese werden dann möglichst vielen Anwendern zur Verfügung gestellt, um die Entscheidungsfindung im Unternehmen zu optimieren. Die leistungsfähige BI-Plattform ist sowohl online als auch offline verfügbar, unterstützt neben Analyse und Reporting auch die Planung und passt sich Ihren Unternehmensanforderungen an.

6.2.1 C8 Importer

Der C8 Importer von Cubeware führt Ihre Daten aus unterschiedlichen Vorsystemen mit dem ETL-Prozess zusammen und stellt gemeinsam mit dem C8 Server das BI Backend dar. Es können verschiedene multidimensionale Datenbanken (SQL-Server, IBM TM1, Infor ION BI, Oracle Hyperion Essbase, OLE DB for OLAP, Microsoft Excel, TXT, CSV, OLE DB/ODBC, MDX und SAP Connect) integriert werden. Diese Fähigkeit hebt das C8 Cockpit von der Microsoft BI Lösung ab. Dieses leistungsfähige Modellierungswerkzeug bringt Ihre unternehmensweiten Daten in eine angemessene Form und hilft Ihnen dabei eine saubere multidimensionale Datenbasis in Ihre BI-Projekte zu bringen. Auch das relationale Datenmanagement wird dank des C8 Importer vereinfacht ermöglicht. Ein ausschlaggebender Vorteil des C8 Importers ist ferner, dass aufgrund der offenen Programmiersprache weitere Applikationen problemlos angebunden werden können. So zeichnet sich der Cubeware Importer durch Schnelligkeit und Einfachheit aus und arbeitet gleichzeitig mit einer sehr hohen Leistungsfähigkeit.

6.2.2 C8 Server

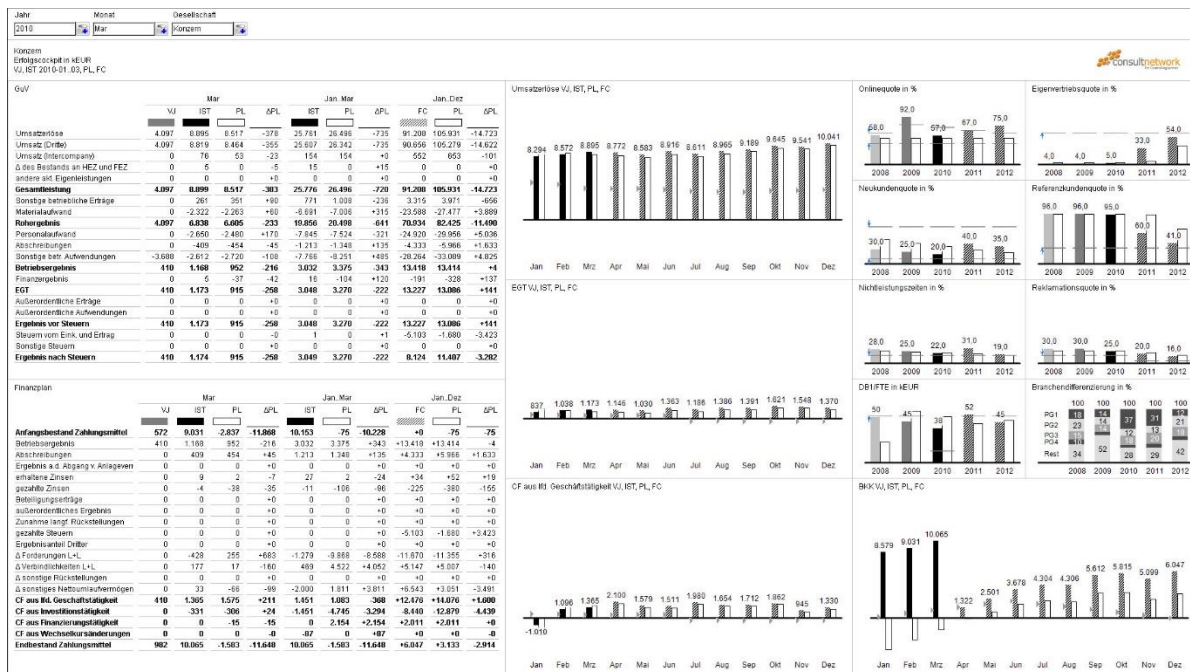
Der C8 Server stellt das Administrationstool dar und ermöglicht ein zentrales Reporting, Daten- und Usermanagement. Er arbeitet mit Planungs-, Reporting- und Analyseprozessen zusammen. Die Reports können automatisch als PDF, PPT, Excel, HTML, XML, usw. zu einer vereinbarten Zeit und an die jeweiligen Empfänger verteilt werden. Mit dem C8 Server ist es möglich, dass unterschiedliche Systeme auf einem Server laufen. Dies funktioniert auf Grund der mehrfach Instanzfähigkeit des C8 Servers. Das bedeutet, dass ein Gerät mehrere Instanzen wie Development, Test oder Production beinhalten kann. Die Instanzen können vom C8 Server sowohl lokal als auch über Clouds verteilt werden. Natürlich kann der C8 Server aber auch auf mehreren Servern installiert werden. Die jeweiligen benötigten Nutzer- und

Berechtigungskonzepte können von Ihnen individuell modifiziert werden. Ihr C8 Server verändert sich mit den Anforderungen Ihres Unternehmens.

6.2.3 C8 Cockpit

Wie bereits im Kapitel 5 (Analyse und Präsentationsschicht) kurz erwähnt, stellt das C8 Cockpit ein leistungsstarkes BI Frontend dar, mit dem Sie Reports im Windows Client, aber auch im Web, erstellen und ansehen können. Das C8 Cockpit ist ein einfach bedienbarer Berichtsdesigner. Das C8 Cockpit bietet nicht nur Interoperabilität zwischen Windows und Web, sondern stellt auch BI-Apps für Android, iOS (iPhone, iPad) und Windows 10 zur Verfügung. Es vereint Analyse, Planung, Dashboarding und das Reporting in einem einzigen Werkzeug. Somit stellt dieses Cockpit ein einheitliches, leicht verständliches und flexibles Planungssystem dar, welches auf die Bedürfnisse aller Abteilungen und Mitarbeiter Ihres Unternehmens abgestimmt ist. Um alle Pläne einheitlich darstellen zu können, werden die relevanten Daten automatisch aus den Vorsystemen herausgefiltert. Die Mitarbeiter erhalten Planungssaplikationen, die auf deren jeweiligen Erfordernisse abgestimmt sind. Damit ist jeder im Unternehmen immer am aktuellsten Stand und Veränderungen sind für alle nachvollziehbar. Somit können Sie Ihre unternehmensweiten Ziele verfolgen, überprüfen, sowie gleichzeitig Zeit und Geld sparen. Wir empfehlen Ihnen deshalb Ihren Erfolg mit der Cubeware Solution Platform C8 zu planen.

Das C8 Cockpit ist ein sehr leistungsfähiges BI-Tool, welches Ihnen eine hohe Dichte an Information liefert. Diese Informationen können mit einem ONE PAGE ONLY (OPO) optimal dargestellt werden. Mit einem OPO Berichtswesen können mehrere Charts auf einer Seite abgebildet werden. Somit werden relevante Sachverhalte für einen Entscheider optimal zusammengefasst. Wenn auch Sie Ihr Reporting auf einen unternehmensweiten Standard bringen wollen, empfehlen wir Ihnen unsere kostenlose dreiteilige Videoserie zum Thema [ONE PAGE ONLY](#), in welcher Sie den Aufbau eines ONE PAGE ONLY Berichtswesens von der Konzeptionierung bis zur technischen Umsetzung erleben können.



Kontakt:

Ing. Mag. Mario Rosenfelder (CEO)
+49 (0)8031 58180-12

+43 (0)463 219095-12

mario.rosenfelder@consultnetwork.com

7 Implementierungsvarianten Data Warehouse

Die Implementierung eines Data Warehouse hängt im Wesentlichen vom zu implementierenden Datenumfang ab. Der Aufbau eines Warehouse kann zum Beispiel schrittweise erfolgen, wobei zwischen zwei Herangehensweisen unterschieden werden kann. Zum einen kann ein umfassendes Data Warehouse implementiert werden, zum anderen gibt es aber auch die Möglichkeit eines bedarfsorientierten DWH.

7.1 Umfassendes DWH

Bei einem umfassenden Data Warehouse werden alle irgendwie relevant erscheinenden Daten, welche im Unternehmen vorhanden sind, in das Warehouse geladen. Hier wird also anfangs nicht darauf geachtet, welche Daten für die verschiedenen verwendeten Applikationen benötigt werden. Erst nachdem die Daten in das Data Warehouse geladen wurden, wird überlegt, wie die Daten verwendet werden.

7.2 Bedarfsorientiertes DWH

Beim bedarfsorientierten Data Warehouse wird vorab die Überlegung gestartet, welche Daten in das Data Warehouse geladen werden sollen. Demnach wird zuerst festgelegt, welche Daten für die verschiedenen Applikationen benötigt werden. Dafür wird ein entsprechendes Datenmodell entwickelt. Erst danach werden die Daten in das Warehouse geladen. Der Aufbau des Data Warehouses orientiert sich also an den Anforderungen der verschiedenen Themengebiete bzw. Applikationen.

8 Zusammenfassung

Ein Data Warehouse stellt eine zentrale Sammelstelle Ihrer internen und externen Daten dar. Der ausschlaggebendste Vorteil eines strukturierten, unternehmensweiten Warehouse ist dabei jener, dass die Daten aus unterschiedlichen Datenquellen bereinigt, integriert und anschließend analysiert werden können. Dies geschieht ohne dass die jeweilige Quelle selbst in ihrer eigentlichen Funktion beeinträchtigt wird. Das Data Warehouse liefert Ihnen eine einheitliche Datenbasis, den sogenannten Single Point of Truth (SPOT). Das Prinzip des SPOTs soll im Falle redundanter Datenbestände einen Datenvorrat hervorbringen, der für Ihr gesamtes Unternehmen Gültigkeit besitzt. Ihr Unternehmen verfügt damit über eine optimierte Datenbank. So können die unternehmensweiten Daten mit verschiedenen Applikationen ausgewertet und Ihre Entscheidungsfindung optimiert werden. Durch das Warehouse wird der unkomplizierte Zugriff auf die Daten aus den verschiedensten Unternehmensbereichen, die zentrale Erstellung und Pflege, sowie die erleichterte Zugriffskontrolle gewährleistet. Zudem können Ihre verdichteten Daten optimal analysiert und präsentiert werden.

Um den gesamten Prozess des Data Warehouse Aufbaus zu realisieren, stehen uns verschiedene leistungsstarke Softwaretools zur Verfügung. So wird der Aufbau zum Beispiel mit den Microsoft Programmlösungen: SQL Server (SSIS und SSAS) oder der Cubeware Programmlösung, dem Cubeware Importer, von uns verwirklicht.

Ist die Implementierung eines unternehmensweiten, zentralen Data Warehouse auch für die Zukunft Ihres Unternehmens erfolgskritisch? Dann kontaktieren Sie uns gleich jetzt und profitieren Sie von den zahlreichen Vorteilen, die durch die Umsetzung eines Data Warehouse entstehen!

CONSULTNETWORK GMBH

Wien | Twin Tower: Wienerbergstraße 11/12a | 1100 Wien

Klagenfurt am Wörthersee | Bahnhofstraße 49 | 9020 Klagenfurt

Rosenheim | Kufsteiner Straße 103 | 83026 Rosenheim

<http://www.controlling-strategy.com>

office@consultnetwork.com

9 Die Autoren

Mario Rosenfelder



Mario Rosenfelder studierte an der Alpen-Adria-Universität Controlling und strategisches Management, machte an der ST. GALLER BUSINESS SCHOOL seine Management Ausbildung für Executives und beschäftigt sich seit über 10 Jahren mit dem Thema Unternehmenssteuerung. Er realisierte und begleitete mit seinem Unternehmen, der Firma consultnetwork Controllingberatung und -dienstleistung GmbH Controllingeinführungsprojekte in über 100 Unternehmen und Unternehmensgruppen. Er ist aktives Mitglied des internationalen Controllervereins, Vortragender in verschiedenen Universitätslehrgängen, Erfinder des FORECAST-KAPAZITÄTS-KONGRUENZ-Modells und Speaker zum Thema integrierte Unternehmenssteuerung. Er veröffentlichte zahlreiche Videobeiträge zu Controlling Themen wie ONE PAGE ONLY-Berichtswesen, Bilanzplanung oder Forecastcontrolling. Er ist der Mastermind hinter der www.controlling-strategy.com und entwickelt diese laufend weiter. Nach wie vor begleitet er TOP-Kunden bei der Umsetzung Ihrer controlling-strategy Projekte.

Bernhard Rosenfelder



beschäftigt sich seit über 10 Jahren mit der Entwicklung von Planungs- und Reportingssystemen auf Basis verschiedener BI-Werkzeuge. Zu seinen Spezialgebieten gehört das Design von umfassenden Data-Warehouse Lösungen. Seit 2008 ist er zertifizierter Cubeware Berater und hat als solcher zahlreiche Projekte mit Cubeware umgesetzt. Neben einer fundierten technischen Ausbildung verfügt er auch über ein abgeschlossenes Betriebswirtschaftsstudium mit dem Schwerpunkt Controlling und Unternehmenssteuerung sowie ein weiteres abgeschlossenes Studium mit dem Schwerpunkt Wirtschaftsrecht. Als Mitgründer und Geschäftsführer der Firma consultnetwork ist er für die Konzeption und Umsetzung der Projekte hauptverantwortlich und Ihr unmittelbarer Ansprechpartner in den Projekten.