

BCBS 239 – Technische Lösungsansätze für Data Warehouse Systeme

integration-factory White Paper

Daniel Feidieker, Michael Schliebitz

12.03.2019

Inhalt

1	Überblick über BCBS 239.....	2
1.1	Hintergrund	2
1.2	Definition.....	2
1.2.1	Gesamtunternehmensführung und Infrastruktur.....	3
1.2.2	Risikodaten-Aggregationskapazitäten.....	3
1.2.3	Risikoberichterstattung.....	4
1.2.4	Aufsichtsrechtliche Überprüfungen, Instrumente und Zusammenarbeit.....	5
1.3	Auswirkungen	6
2	Technische Lösungsansätze.....	7
2.1	Architektur / Layer Model / Datenmodell.....	7
2.2	Metadaten gestützte Template-Entwicklung	8
2.3	Bitemporale Modellierung.....	10
2.4	Data Vault-Modellierungsansatz	12
3	Über integration-factory	16

Abbildungsverzeichnis

Abbildung 1-1:	BCBS 239 Clusterung Grundsätze.....	2
Abbildung 2-1:	Layer Model.....	7
Abbildung 2-2:	Design-Template-Aufbau	9
Abbildung 2-3:	Darstellung Historisierungstypen.....	10
Abbildung 2-4:	Zeitintervalle Bi-Temporale Darstellung	11
Abbildung 2-5:	Data Vault Modellierungsansatz.....	13
Abbildung 2-6:	Darstellung Ladeprozess	13
Abbildung 2-7:	Architektur – Data Vault.....	14
Abbildung 2-8:	Architektur – Klassisch.....	14
Abbildung 2-9:	Architektur – Raw / Business Vault.....	15

1 Überblick über BCBS 239

1.1 Hintergrund

Der Niedergang von Lehmann Brothers und die anschließende globale Finanzkrise haben gezeigt, dass viele Banken keine adäquate Risikosteuerung und insbesondere keine geeigneten IT Systeme für das Risikomanagement zur Verfügung hatten. Häufig lagen Daten dezentral im Unternehmen vor und konnten nicht oder nur schwer über Abteilungsgrenzen oder Konzerngesellschaften hinweg zusammengeführt werden. Ein unternehmensweites Reporting von Kennzahlen, Risiken, etc. war so nicht möglich. Der Aufwand zur Ermittlung des tatsächlichen Risikos war somit hoch und zeitintensiv und führte dazu, dass Entscheidungen ohne echte analytische Grundlage getroffen wurden.

Um in Zukunft schnell und vor allem rechtzeitig eine Indikation über den Zustand eines Finanzinstituts erhalten und besser auf äußere Ereignisse reagieren zu können, verabschiedete der Baseler Ausschuss für Bankenaufsicht (Basel Committee on Banking Supervision, kurz BCBS) die „Grundsätze für die effektive Aggregation von Risikodaten und die Risikoberichterstattung“ (BIS)¹, die in der Leitlinie BCBS 239 veröffentlicht wurden.

1.2 Definition

Die in der Leitlinie BCBS 239 enthaltenen Grundsätze lassen sich in vier Themenbereiche gliedern:

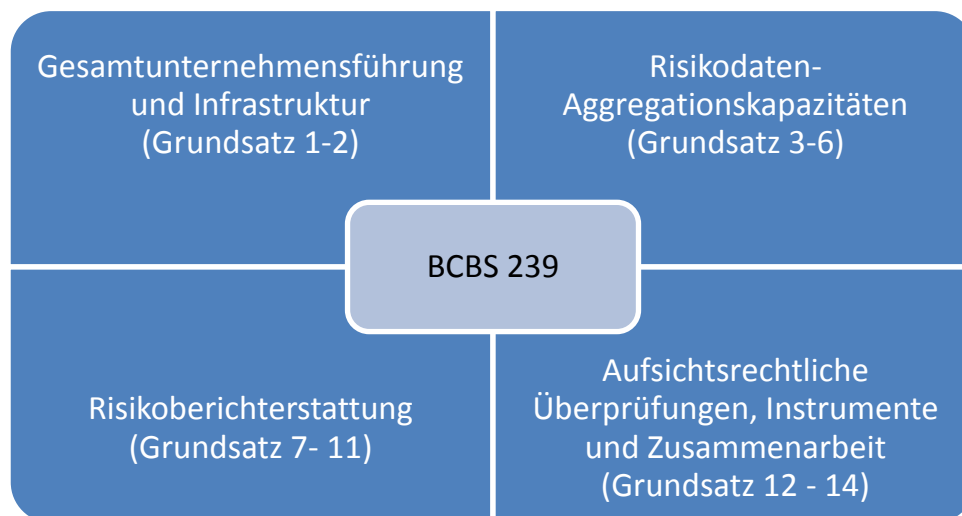


Abbildung 1-1: BCBS 239 Clusterung Grundsätze

¹ Bank for International Settlements, BCBS 239, Online-Publikation, bis.org/publ/bcbs239_de.pdf, Stand: 12. September 2016

1.2.1 Gesamtunternehmensführung und Infrastruktur

Innerhalb des Themenblocks **Gesamtunternehmensführung und Infrastruktur** werden alle Handlungen zusammengefasst, die sich mit dem Aufbau einer stabilen IT-Infrastruktur sowie der darin enthaltenen Daten befassen. Hiermit soll ein Instrument für die Unternehmensführung geschaffen werden, um die Umsetzung der Grundsätze sowie die Meldung an die Aufsichtsinstanzen überwachen zu können.

Grundsatz 1: Governance

Die Prozesse der Unternehmensführung sollen unter Beachtung aller Grundsätze und Empfehlungen, die vom Baseler Ausschuss veröffentlicht wurden, definiert und befolgt werden. Die Unternehmensführung ist dazu verpflichtet, die Regelungen zu überprüfen und für deren Einhaltung zu sorgen. Hierfür sind ausreichende Ressourcen zur Verfügung zu stellen.

Grundsatz 2: Datenarchitektur und IT-Infrastruktur

Unter dem Grundsatz 2 ist der Aufbau einer IT-Infrastruktur innerhalb eines Finanzdienstleistungsinstituts zu verstehen, die zu jeder Zeit, auch in Krisenzeiten, die Risikoberichterstattung unterstützt und gewährleistet. Entsprechende Maßnahmen sollen im Business Continuity Management hinterlegt werden. Zu den Bestandteilen einer soliden IT-Infrastruktur gehören u. a. das Festlegen von Aufgaben und Verantwortlichkeiten, die Erstellung einer konzernweiten Datenarchitektur und die Definition von Datentaxonomien zur Datenharmonisierung.

1.2.2 Risikodaten-Aggregationskapazitäten

Der Auf- und Ausbau der **Risikodaten-Aggregationskapazitäten** muss in der Weise gestaltet werden, dass die Risikomanagementberichte die Unternehmensrisiken glaubwürdig widerspiegeln. Dieser Punkt darf nicht zulasten einer anderen Anforderung vernachlässigt werden.

Grundsatz 3: Genauigkeit und Integrität

Eine genaue, fehlerfreie Aggregation der Risikodaten muss durch eine Bank gewährleistet sein, um den Ansprüchen des Berichtswesens in jeder Marktphase gerecht zu werden. Idealerweise werden die Risikodaten automatisch verarbeitet, um manuelle Fehler zu minimieren. Das Prüfen der Daten, z.B. durch Abgleich mit externen Datenquellen, stellt einen weiteren Baustein dar.

Grundsatz 4: Vollständigkeit

Um dem Grundsatz Vollständigkeit zu genügen sind innerhalb eines Finanzdienstleistungsunternehmens alle Risikodaten des gesamten Konzerns zu aggregieren. Die Aggregation kann hierbei in relevanten Kategorien (bspw. nach Geschäftsfeldern, Branche, Region, ...) erfolgen.

Grundsatz 5: Aktualität

Die Bereitstellung der aggregierten aktuellen Risikodaten muss in einem angemessenen zeitlichen Rahmen liegen. Der Zeitpunkt ist dabei vom zugrundeliegenden Risiko und dessen Volatilität abhängig. Weiteren Einfluss auf die Erstellungsperiode der Risikodaten bilden interne Anforderungen zum Erstellungszeitpunkt sowie der Anteil am gesamten Risikoprofil der Bank.

Grundsatz 6: Anpassungsfähigkeit

Die Reporting-Umgebung einer Bank muss die Generierung einer Vielzahl von Ad hoc-Berichten gewährleisten, wie es bspw. in Stressphasen oder Krisen nötig ist, um Aufsichtsinstanzen mit den geforderten Risikokennzahlen zu versorgen. Hierfür ist eine flexible Anpassung von Prozessen, Berichten, Daten, usw. erforderlich, um den Anforderungen der Aufsichtsinstanzen nachkommen zu können

1.2.3 Risikoberichterstattung

Die **Risikoberichterstattung** stellt sicher, dass dem Management die richtigen Daten zum richtigen Zeitpunkt vorliegen, um ein ordnungsgemäßes Risikomanagement betreiben zu können. Die Berichte müssen dabei genau, klar und umfassend sein.

Grundsatz 7: Genauigkeit

Die in Risikomanagementberichten enthaltenen Risikodaten müssen exakt sein und Risiken wahrheitsgemäß widerspiegeln, um der Geschäftsführung bei Risikoentscheidungen ein genaues Bild zu geben. Damit die Genauigkeit gewährleistet werden kann, bedarf es der Etablierung von Prozessen zum Abgleich von Reports und der darin enthaltenen Daten.

Grundsatz 8: Umfassender Charakter

Die Risikomanagementberichte müssen alle Risikobereiche einer Bank umfassen. Der Detaillierungsgrad ist dabei an die Struktur und Bedeutung der Bank, dem Risikoprofil sowie an die Ansprüche des Empfängers des Risikomanagementberichts anzupassen. Inhalte eines solchen Berichtes können daher Informationen aus den Bereichen Liquiditäts-, Markt- oder Kreditrisiko sein und relevante Risikokennzahlen wie das Eigenkapital ausweisen.

Grundsatz 9: Klarheit und Nutzen

Der Inhalt von Risikomanagementberichten ist auf die Relevanz und den Nutzen für den Empfänger abzustimmen. Um objektive Entscheidungen zu ermöglichen, sollten Berichte leicht verständlich aber auch lückenlos sein. Durch das Einholen eines Feedbacks in bestimmten Intervallen wird die adressatengerechte Aufbereitung sichergestellt.

Grundsatz 10: Häufigkeit

Die Intervalle zur Generierung und dem Versenden von Risikomanagementberichten werden von der Unternehmensleitung bestimmt. Diese sind von den Interessen der Empfänger, der Risikoart, der Volatilität von Risiken sowie der Wichtigkeit der Berichte für ein ordnungsgemäßes Risikomanagement abhängig.

Grundsatz 11: Verbreitung

Bei der Verbreitung der Risikomanagementberichte an den Adressatenkreis sind entsprechende Prozesse zur Übermittlung und zur Wahrung der Vertraulichkeit zu etablieren. Ferner wird die regelmäßige Überprüfung des ordnungsgemäßen Berichteingangs bei den Adressaten durch die Bank von den Aufsichtsbehörden erwartet.

1.2.4 Aufsichtsrechtliche Überprüfungen, Instrumente und Zusammenarbeit

Der Punkt **Aufsichtsrechtliche Überprüfungen, Instrumente und Zusammenarbeit** beschreibt die Rolle und Aufgaben der Aufsichtsinstanzen zur Überprüfung der zuvor genannten Grundsätze.

Grundsatz 12: Überprüfung

Die zuvor erwähnten Grundsätze werden durch die Aufsichtsbehörden in regelmäßigen Abständen kontrolliert. Die Überprüfung findet nicht nur während des regulären Betriebs statt, sondern erfolgt ebenso in Krisenzeiten. Bei der Überprüfung ist den Aufsichtsinstanzen z. B. Zugang zu allen wesentlichen Dokumenten zu gewähren sowie Raum für den persönlichen Austausch mit Personen wie externen Revisoren oder von der Bank unabhängigen Experten einzuräumen.

Grundsatz 13: Korrektur- und Aufsichtsmaßnahmen

Damit Aufsichtsbehörden Korrekturen von Mängeln in der Risikoberichtserstattung bei Banken einfordern können, müssen sie über ausreichende Instrumente und Ressourcen verfügen. Hierzu gehören beispielsweise die enge Überwachung, das Anordnen zur Behebung von Korrekturen durch eine Bank sowie die Anweisung einer Prüfung durch eine unabhängige dritte Partei.

Grundsatz 14: Grenzüberschreitende Zusammenarbeit

Bei der Überprüfung der Grundsätze sowie Realisierung von Korrekturen müssen nationale Aufsichtsbehörden mit Instanzen anderer Länder zusammenarbeiten. Durch diesen grenzüberschreitenden Austausch wird die Überprüfung aller Geschäftsaktivitäten einer Bank in mehreren Ländern gefördert. Ebenso lassen sich durch diesen Austausch gleiche Prüfungen unterschiedlicher Instanzen vermeiden.

1.3 Auswirkungen

Die durch BCBS 239 aufgestellten Grundsätze haben Auswirkungen auf die IT-Systeme der Unternehmen aus der Finanzbranche. Eine stress- und krisenfeste Architektur des Gesamtsystems stellt die Basis aller weiteren Maßnahmen dar.

Die Systeme müssen eine schnelle und umfassende Berichtserstellung unterstützen, um geforderte Kennzahlen den Aufsichtsbehörden zur Verfügung zu stellen. Um auch auf kurzfristige außerplanmäßige Anfragen reagieren zu können, müssen die IT-Systeme flexible Auswertungsmöglichkeiten unterstützen (Ad hoc Reporting). Ein weiterer wichtiger Baustein ist das Datenqualitätsmanagement, das Funktionen für die Messung der Datenqualität bereitstellt.

Konkret können diese Anforderungen und im Speziellen insbesondere die Grundsätze 2, 3, 4 und 5 mit einem Financial Data Warehouse als integrierte und harmonisierte Plattform im Unternehmen umgesetzt werden. Durch die umfassende Integration von relevanten Daten aus allen Unternehmensbereichen entsteht ein zentrales Data Warehouse für die Unternehmenssteuerung. Werden bei der Datenintegration die Daten konsolidiert und in das normalisierte Datenmodell gespeichert, unterstützt dies die Vergleichbarkeit und die kontextgleiche Auswertung von zuvor nicht verknüpfbarer Daten. Eine Analyse ist hierbei nicht nur auf dem aktuellen Stand der Daten möglich (as-is), sondern es wird auch die Abfrage historischer Daten gewährleistet (as-was). Das Data Warehouse ist damit in der Lage, jederzeit Auskunft über die Entwicklung des Datenbestandes zu geben und sichert infolgedessen die vollständige Nachvollziehbarkeit des Datenhaushalts ab.

2 Technische Lösungsansätze

2.1 Architektur / Layer Model / Datenmodell

Zur Schaffung eines zentralen und normalisierten Data Warehouse, das alle Daten für die Unternehmenssteuerung bereitstellt, ist die Nutzung eines Layer Modells innerhalb des Data Warehouse unumgänglich. Mithilfe verschiedener Layer lassen sich strukturunterschiedliche Quelldaten in ein normalisiertes Datenmodell überführen, bevor die Daten anschließend einzelnen Abnehmern zur Verfügung gestellt werden können. Das folgende Schaubild stellt einen möglichen Aufbau dar:

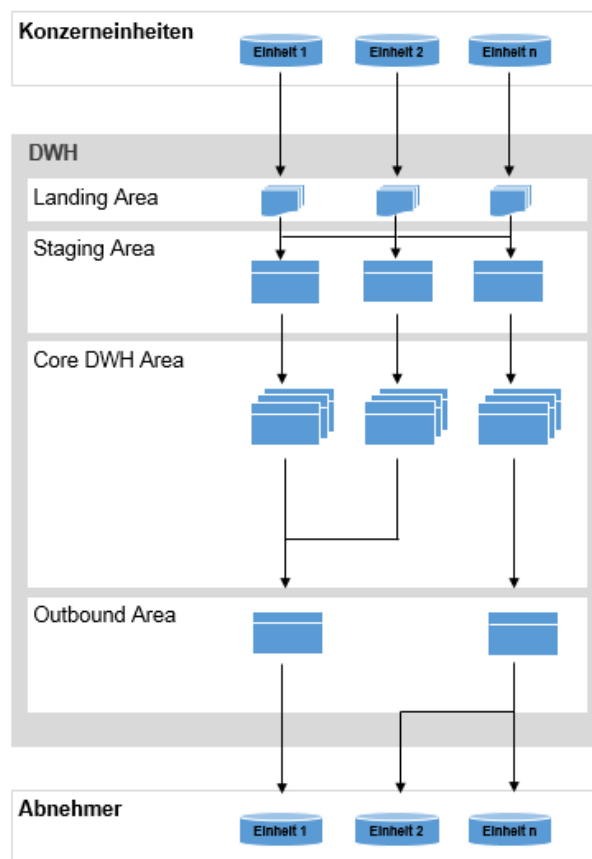


Abbildung 2-1: Layer Model

Landing Area

Der Landing Bereich stellt die erste Stufe des DWH dar. In diesem Layer werden alle Lieferobjekte in eine relationale Form überführt und in separaten Landing-Tabellen vorgehalten. Beim Datenimport wird die Quellstruktur 1:1 übernommen und auch Datentypkonvertierungen finden an dieser Stelle noch nicht statt.

Staging Area

Innerhalb der Staging Area findet die Transformation der Quellstrukturen in die normalisierte Sicht des Unternehmensdatenmodells statt. Diese Schicht beinhaltet den Großteil der fachlichen Transformationslogik, welche von der reinen Datentypkonvertierung bis hin zu komplexen fachlichen

Abbildungen reicht. Mittels Einsatz verschiedener Logging-Komponenten können auftretende Fehler bzw. Inkonsistenzen individuell protokolliert werden.

Core DWH Area

Die zuvor gebildete DWH-Sicht aus den Quellstrukturen heraus wird bei der Bewirtschaftung des Core DWH historisiert. Somit lassen sich alle bisher aufgetretenen Zustandsänderungen im Lebenszyklus eines Objektes jederzeit nachvollziehen und auswerten. Weiterhin werden technische Surrogate Keys auf Basis des fachlichen Schlüssels pro Fachobjekt gebildet, um im weiteren Verlauf Beziehungen zwischen den Objekten verwalten zu können.

Dieser Layer-Ansatz unterstützt den BCBS 239 Themenbereich Gesamtunternehmensführung und Infrastruktur, im Speziellen den Grundsatz 2 Datenarchitektur und IT-Infrastruktur.

2.2 Metadaten gestützte Template-Entwicklung

Zur Unterstützung einer effizienten, metadatengestützten Entwicklung in Data Warehouse-Projekten hat integration-factory Datenintegrationstemplates entwickelt, um gleichartige Prozesse in gleichbleibender Funktionalität und Qualität umsetzen zu können. Diese Templates stellen in ihrer Urform die Historisierung von Stamm- und Bewegungsdaten zur Nachvollziehbarkeit der Datenbewegungen in einem Data Warehouse dar. Darüber hinaus verfügen die Templates über einen Sicherungsmechanismus der unveränderten Eingangsdaten, um eine Nachvollziehbarkeit und Dokumentierung zu gewährleisten. Zur Erhöhung der Datenqualität im Core-Warehouse bieten die Templates weiterhin die Möglichkeit, spezifische Prüfregele für die Datenverarbeitung hinterlegen zu können. Je nach Erfüllungsgrad werden die Daten automatisiert ins Core-Warehouse (= valide Daten) oder in einen speziellen Reject-Container (= nicht valide Daten) gespeichert. Der Reject-Container ermöglicht die Analyse und spätere Nachverarbeitung der abgelehnten Daten. Grundsätzlich wird innerhalb der Templates der „Insert-Only“-Ansatz verfolgt, um eine vollständige Nachvollziehbarkeit der Datenhaltung erzielen zu können.

Die Methodik der Templates ist im Wesentlichen softwareunabhängig und kann gemäß den Projektgegebenheiten für Tools wie beispielsweise Informatica PowerCenter, Talend DataIntegration oder Microsoft SQL Server Integration Services bereitgestellt werden. Die eingesetzten Templates sind damit anwendungskontextneutral und können für unterschiedlichste Zwecke und Branchen eingesetzt werden. Durch den Einsatz standardisierter Komponenten wird weiterhin die gleichbleibende Verarbeitungslogik sichergestellt. Einmal implementiert und ausführlich getestet gewährleistet der Einsatz von Templates eine Effizienzsteigerung in der Entwicklungsphase sowie eine gleichbleibende Implementierungsqualität bei der Vervielfältigung für unterschiedliche Einsatzgebiete.

Die Verwendung von Templates innerhalb der Datenintegration decken Grundsätze unterschiedlicher Themenbereiche ab bzw. schaffen Grundlagen, um die Grundsätze in nachgelagerten Prozessen zu erfüllen. Der Einsatz standardisierter Templates ist beispielsweise ein Baustein für den Aufbau einer IT-Infrastruktur mit robusten Datenintegrationsprozessen, wie es im Grundsatz 2 des Themenbereichs Gesamtunternehmensführung und Infrastruktur beschrieben wird. Weiterhin stellen Datenintegrationsprozesse eine Möglichkeit dar, um Prüfregele und Logging-Informationen standardisiert zu hinterlegen, bevor diese anschließend in einem nachgelagerten Prozess/ System ausgewertet und für das übergeordnete Datenqualitätsmanagementsystem aufbereitet werden. Standardisierte Templates stellen damit eine Maßnahme dar, um die vollständige und korrekte Datenhaltung für die Risikodatenaggregation zu gewährleisten.

Templates für die standardisierte Verarbeitung wurden für den Bereich Stammdaten, Beziehungen unter Stammdaten sowie Fakten/ Bewegungsdaten entwickelt. Die Templates bestehen aus Datenstrukturen und Datenintegrationsprozessen als Algorithmus zur Bewirtschaftung dieser Strukturen. Abbildung 2-2 stellt schematisch die Beziehung der drei Entitätsgruppen zueinander sowie das jeweilige Datenmodell und das Prozesspattern anhand des Datenintegrationstools Informatica PowerCenter Designer dar.

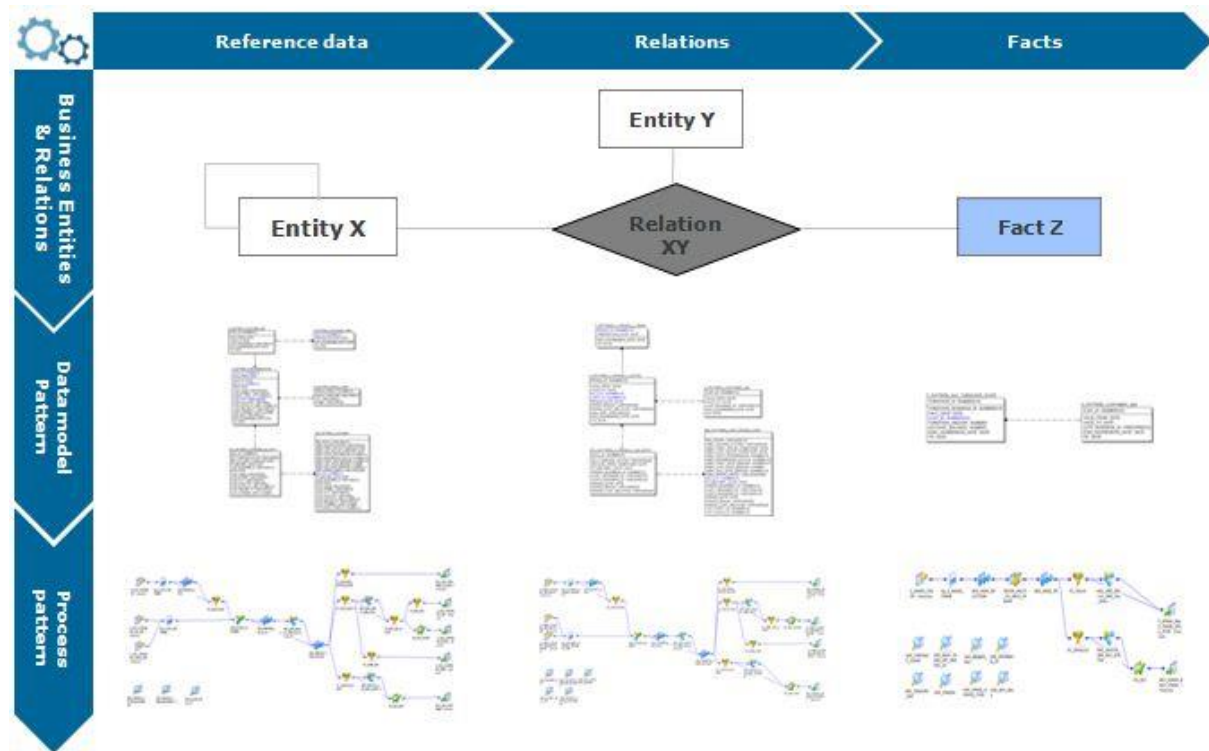


Abbildung 2-2: Design-Template-Aufbau

Aufgrund differenzierter Anforderungen an die einzelnen Entitätsgruppen bieten die Datenmodelle und Prozesstemplates unterschiedliche Komplexitätsgrade. Allgemeingültige Standards lassen sich dennoch überall wiederfinden. Beispielsweise folgen die vorgesehenen relationalen Datenmodelle der 3. Normalform, um Redundanzen in der Datenhaltung zu vermeiden.

Weiterhin enthält das Datenmodell für Referenzdaten einen Surrogate Key, mit dem Realwelt-Objekte eindeutig referenziert werden können. Dieser besitzt eine lebenslange Gültigkeit. Zusätzlich weist jedes Realwelt-Objekt einen Alternativ-Schlüssel auf, der sich aus dem Business Key als Database Constraint (Unique Key) definieren lässt.

Die im zeitlichen Kontext auftretenden historischen Zustände von Objekten werden normalisiert in eigenen Zustandstabellen mit den Intervallgültigkeitsgrenzen `VALID_FROM` und `VALID_TO` gespeichert. `VALID_TO` ist eine Information, die sich aus der Chronologie der Objektzustände gemäß `VALID_FROM` und expliziter Terminierungen als `Nachfolgezustand-VALID_FROM - 1` ergibt (z.B. minus eine Sekunde).

Transaktionsbasierte Daten, also Fakten, werden im Gegensatz zu den Referenzdaten zeitpunktbezogen verwaltet. Die Speicherung erfolgt entsprechend mit einem fachbezogenem Datum `FACT_DATE` sowie einem technisch orientierten Datum `AWARENESS_DATE`, welches den Zeitpunkt der Speicherung im Core Data Warehouse widerspiegelt. Über das `AWARENESS_DATE`

lassen sich effizient in Aggregationsprozessen die noch nicht berücksichtigten Informationen ermitteln.

Bei der Faktenverarbeitung wird ebenfalls der „Insert-Only“-Ansatz verfolgt, da jedes Delete oder Update Informationsverlust bedeuten würde und ggf. dazu führt, dass ausgelieferte Ergebnisse nicht mehr reproduzierbar wären. Löschungen/ Stornierungen werden daher als Gegenbuchung bzw. neue Version gespeichert. Bei einer Abfrage ist entsprechend immer die letzte Version zu einem bestimmten Zeitpunkt zu wählen.

Die zuvor erläuterten intervall- und zeitpunktbezogenen Historisierungsarten greift die nachfolgende Abbildung auf:

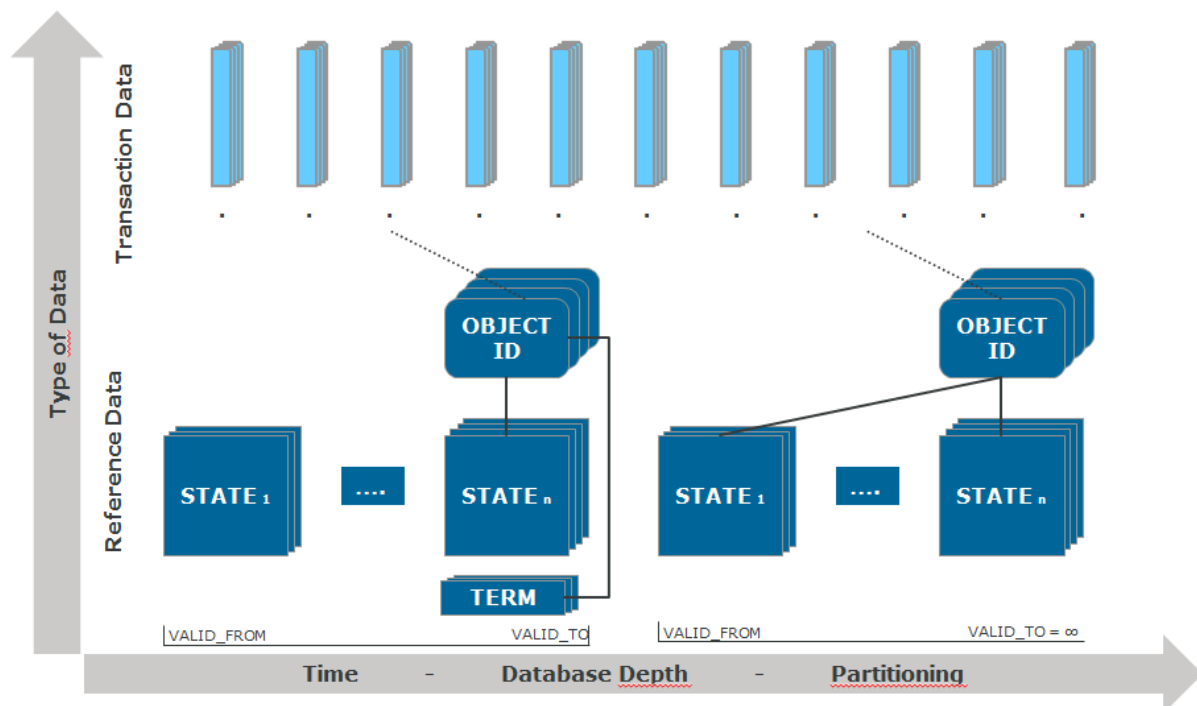


Abbildung 2-3: Darstellung Historisierungstypen

2.3 Bitemporale Modellierung

Eine der Grundaufgaben eines jeden Data Warehouses ist im Gegensatz zur Zeitscheiben-Datenhaltung die zeitliche Historisierung von Unternehmensdaten (= temporale Datenhaltung), die als Ausgangsbasis für die Verwendung in Berichten oder anderen nachgelagerten Systemen zur Verfügung stehen müssen. Es reicht meist nicht aus, nur die aktuelle Sicht auf die Daten vorzuhalten, welche bei auftretenden Änderungen die im Data Warehouse vorgehaltenen Daten überschreiben. Stattdessen ist aus Revisionsgründen eine Nachvollziehbarkeit gewünscht, weshalb das Vorhalten aller Datenänderungen erforderlich ist. Dies ermöglicht aber auch historische Datenbestandsanalysen mit darauf aufbauenden Forecasts.

Um dies zu ermöglichen existieren in der temporalen Modellierung zwei Arten, wie Daten zeitlich eingeordnet werden können:

Gültigkeitszeitraum:

- Abbildung der fachlichen Gültigkeit der Attribute
- Abbildung des Zeitraums mittels Gültig-Ab und Gültig-Bis, Zeitpunkt fachlicher Stichtag
- Vorgabe muss durch den Benutzer erfolgen

Transaktionszeit:

- Abbildung der technischen Gültigkeit der Attribute (Zeitpunkt der Speicherung im Data Warehouse)
- Zeitprojektion kann hierbei nicht in der Zukunft liegen
- Vorgabe kann vom System erfolgen

Finden beide Zeitdarstellungen Anwendung, so wird von einem bi-temporalen Modellierungsansatz gesprochen.

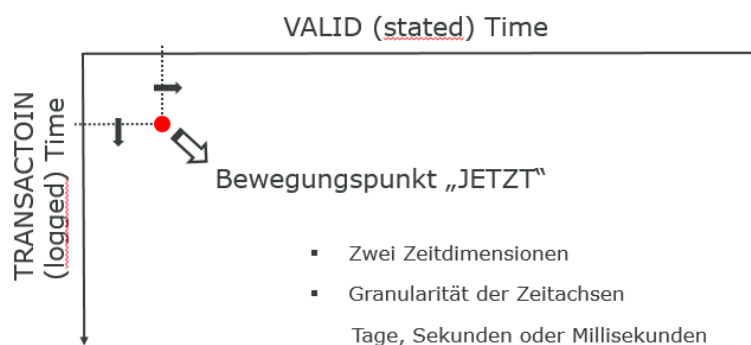


Abbildung 2-4: Zeitintervalle Bi-Temporale Darstellung

Wie bereits erwähnt wird mit dieser Modellierung die Revisionsicherheit gewährleistet. Ebenso lassen sich Abfragen bei der Weiterverwendung der Daten bzw. im Reporting zu einem beliebigen Zeitpunkt jederzeit reproduzieren (AS OF-Analysen / Abfragen).

Das nachfolgende Beispiel zeigt die Funktionsweise der bi-temporalen Modellierung mit dem Insert-Only-Gedanken auf und wie sich im zeitlichen Verlauf die Datenhistorie aufbaut.

Die Person Max Mustermann lebt seit dem 01. Januar 2014 in Hamburg. Dieser Sachverhalt wurde am 04.01.2016 im Data Warehouse gespeichert.

NAME	Position	VT_START	VT_END	TT_START	TT_END
Max Musterman	Consultant	01.01.2014	∞	04.01.2016	∞

Am 10.01.2016 wurde festgestellt, dass sich beim Nachnamen ein Fehler in der Schreibweise eingeschlichen hat. Der korrigierte Datensatz wird daraufhin im Data Warehouse neu abgelegt, der Transaktionszeitpunkt des bereits existierenden Datensatzes wird aktualisiert.

NAME	Position	VT_START	VT_END	TT_START	TT_END
Max Musterman	Consultant	01.01.2014	∞	04.01.2016	∞ 09.01.2016
Max Mustermann	Consultant	01.01.2014	∞	10.01.2016	∞

Max Mustermann wurde zum 01.06.2016 zum Senior Consultant befördert, im Data Warehouse wird diese Beförderung am 04.06.2016 verarbeitet. Hieraus entsteht eine neue gültige Version, die ab dem 01.06.2016 gilt. Die zuvor gültige Version wird zunächst aus technischer Sicht abgeschlossen und als weiterer Datensatz erneut im Data Warehouse eingebucht, um die Terminierung aus fachlicher Sicht zum 31.05.2016 darstellen zu können.

NAME	Position	VT_START	VT_END	TT_START	TT_END
Max Musterman	Consultant	01.01.2014	∞	04.01.2016	09.01.2016
Max Mustermann	Consultant	01.01.2014	∞	10.01.2016	03.06.2016
Max Mustermann	Consultant	01.01.2014	31.05.2016	04.01.2016	∞
Max Mustermann	Snr. Consultant	01.06.2016	∞	04.06.2016	∞

Wie man sieht, ermöglicht diese Historisierungsform die Auswertung in jede beliebige Richtung. Eine „Most-Recent“-Sicht, die Max Mustermann heute als Senior Consultant ausweist, ist genauso möglich wie die Frage, welche Position Max Mustermann im Jahr 2015 bekleidete. Der bisherige Karriereverlauf von Max Mustermann aus heutiger Sicht ergibt, dass er von Januar 2014 bis Mai 2016 als Consultant und seit dem Juni 2016 als Senior Consultant tätig ist. Hierbei kann es sich auch um eine immer wiederkehrende Abfrage handeln, die je nach Abfragezeitpunkt unterschiedliche Ergebnisse liefert. Ist es aber notwendig, das Ergebnis zum Abfragezeitpunkt 01.05.2016 zu reproduzieren, so ist dies durch das Abändern des Transaktionszeitpunkts machbar. Das Ergebnis weist Max Mustermann dann als Consultant aus.

2.4 Data Vault-Modellierungsansatz

Eine Methodik, die für den Aufbau einer Datenarchitektur und damit einhergehend zur Sicherstellung des Grundsatzes 1 verwendet werden kann, ist der Data Vault-Modellierungsansatz. Bei Data Vault handelt es sich um eine Data Warehouse Methodensammlung, die den Aufbau eines Enterprise Data Warehouse verfolgt. Entwickelt wurde sie von Dan Linstedt (1990-2000) und stellt einen aktuellen Trend im Data Warehouse Umfeld dar.

Im Bereich der Datenmodellierung weist Data Vault konzeptionelle Elemente der Modellierung auf, gepaart mit dem Hub & Spoke-Architekturansatz sowie der dimensionalen Modellierung. Im Gegensatz zu anderen Modellierungsansätzen wie dem Star Schema (Dimensionale Modellierung) oder 3NF (Normalisierte Modellierung) werden die fachlichen Elemente in ihre Bestandteile aufgeteilt.

Kernelemente des Data Vault Modells stellen die HUBs, LINKs und SATelliten dar. Der HUB steht für ein Objekt im Business Kontext und enthält die fachlichen Schlüssel (Business Keys). Diese sind nahezu unveränderlich und tragen so zur Stabilität des Modells bei. Innerhalb der LINKs werden die Beziehungen der Objekte/ Business Keys zueinander abgebildet. Die SATs speichern die abhängigen Attribute der HUBs und LINKs.

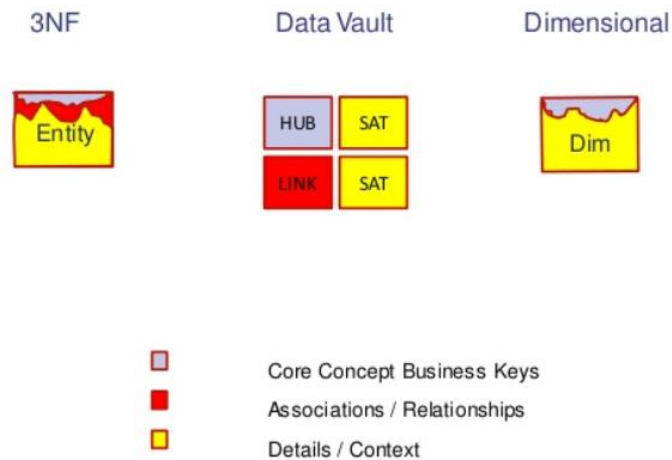


Abbildung 2-5: Data Vault Modellierungsansatz2

Durch diese Aufteilung wird mit dem Data Vault Ansatz ein jederzeit anpassbares, skalierbares Data Warehouse geschaffen, mit dem schnell auf Veränderungen reagiert werden kann, ohne bestehende Strukturen und Prozesse abändern zu müssen. Soll ein fachliches Objekt neue Attribute aus einem weiteren Quellsystem erhalten, können diese Attribute in einen neuen Satelliten ausgelagert werden. Die Erstellung des neuen physischen Satelliten mit den erforderlichen Bewirtschaftungsprozessen kann parallel zur existierenden Welt erfolgen, ohne diese verändern oder anpassen zu müssen.

Data Vault bietet weiterhin die Möglichkeit, im Entwicklungsprozess Verbesserungspotenziale zu erzielen. Durch die Reduzierung auf drei Modellbestandteile wird ein hoher Standardisierungsgrad erreicht, der den Einsatz von Bewirtschaftungstemplates und Generatoransätzen ermöglicht. Die im Kapitel 2.2 erwähnten Templates können auf die Data Vault Modellierung adaptiert werden.

Die Methoden der Datenverarbeitung unterstützen hingegen die Standardisierung des Datenintegrationslayers. Dies führt u.a. zur Reduzierung komplexer Jobnetze, da jeweils HUBs, LINKs und SATs parallel zueinander bewirtschaftet werden können.

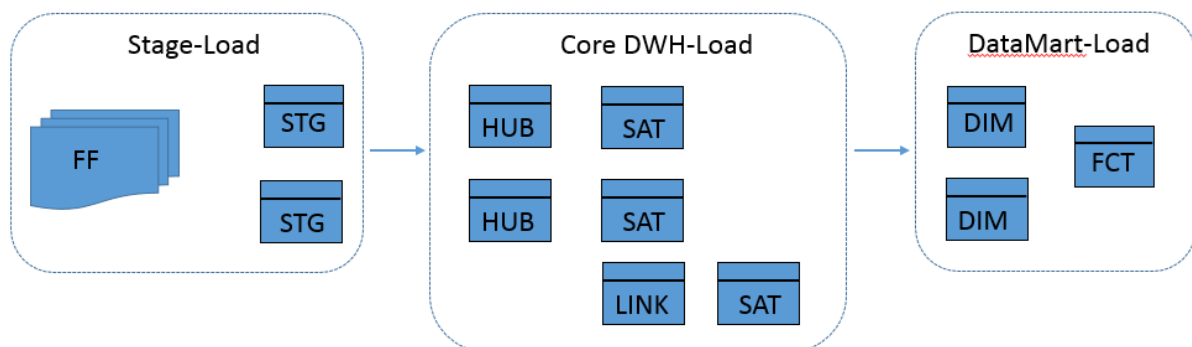


Abbildung 2-6: Darstellung Ladeprozess

² Data Vault Modellierungsansatz, Hans Hultgren, <http://de.slideshare.net/HansHultgren/2014-ensemble-modelingpost>

Steht die Bewirtschaftung des Core Data Warehouse an, werden zunächst alle HUB-Prozesse gestartet, um die eigenständigen Objekte mit ihren fachlichen Schlüsseln zu verwalten. Anschließend erfolgt die Bewirtschaftung der LINKs sowie derjenigen SATs, die einem HUB zugewiesen sind. Die zu einem LINK gehörenden SATs werden aufbereitet, sobald die erfolgreiche Verarbeitung der LINK Informationen abgelaufen ist. Durch diese Reduzierung der Abhängigkeiten ist eine hoch parallellaufende Verarbeitung möglich, lediglich eingeschränkt durch wirklich notwendige Abhängigkeiten sowie der zugrundeliegenden Hardware und der Prozesslastverteilung.

Die Verarbeitung bei Data Vault ist im einfachen Architekturansatz davon geprägt, Quelldaten ohne große fachliche Transformationslogik im Data Warehouse zu integrieren. Fachlich notwendige Logik wird erst anschließend auf den Datenbestand angewendet, um die Ergebnisse in nachgelagerten Systemen, Data Marts, Reports, etc. zu verwenden.

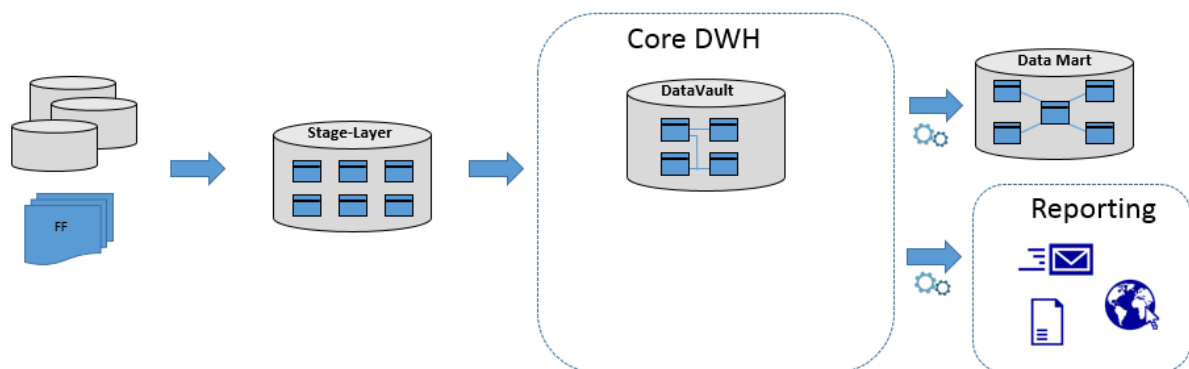


Abbildung 2-7: Architektur – Data Vault

Dieser Ansatz stellt das Gegenteil der klassischen Data Warehouse Architektur dar. Diese zeichnet sich durch die Anwendung der fachlichen Transformationslogik bei der Befüllung des Core Data Warehouse aus. Die Daten im DWH entsprechen somit dem fachlichen Zielbild, eine redundante Implementierung der Transformationslogik wie beim einfachen Data Vault Ansatz ist nicht notwendig.

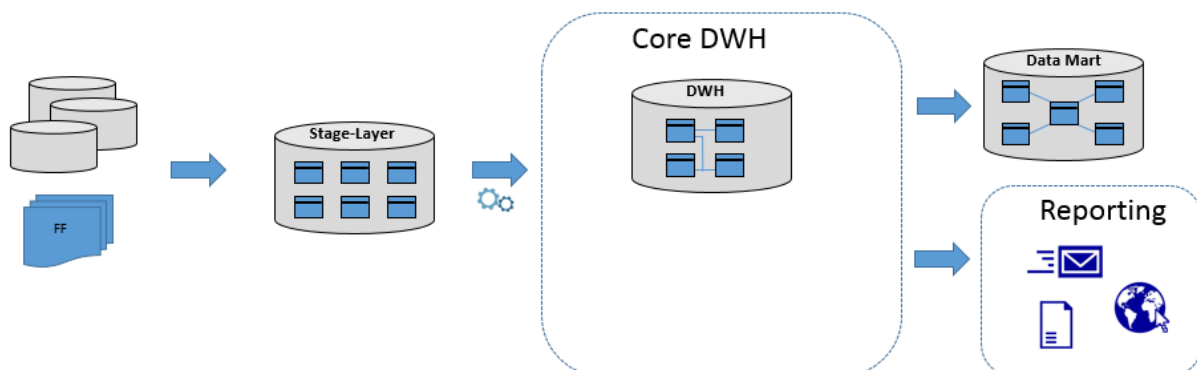


Abbildung 2-8: Architektur – Klassisch

Eine Mischform beider Architekturansätze stellt die Abbildung 2-9 vor. In einem ersten Schritt wird der sogenannte Raw DataVault mit den Quellinformationen ohne Transformationslogik befüllt. Auf dieser Datenbasis erfolgt im Nachgang die Anwendung fachlicher Mappinglogik um den Business Vault erstellen zu können. Der Business Vault stellt wie im klassischen Architekturansatz die fachliche Sicht innerhalb des Data Warehouse dar. Durch den Aufbau eines Raw- sowie Business Vaults besteht die Möglichkeit, auf die Ergebnisse der fachlichen Transformationslogik (Business Vault-Sicht) bzw. auf die originären Quellinformationen im Raw DataVault zugreifen zu können.

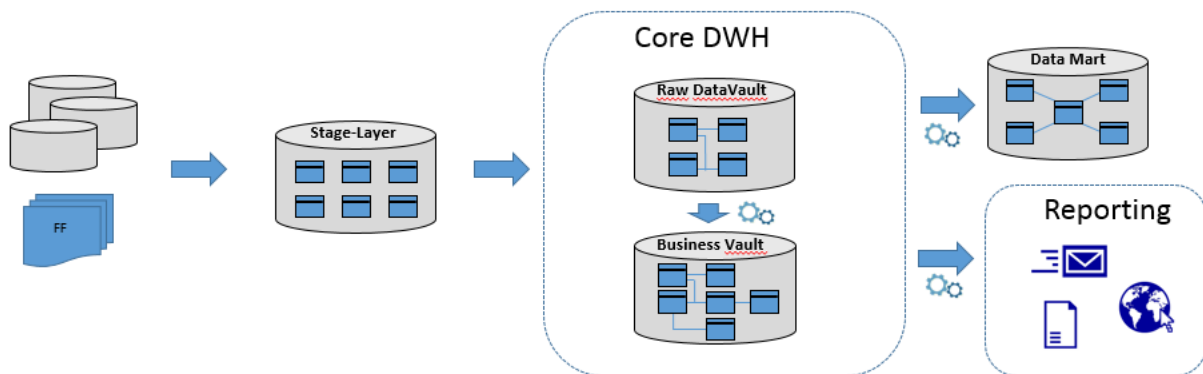


Abbildung 2-9: Architektur – Raw / Business Vault

Data Vault revolutioniert nicht die Data Warehouse Architektur, Modellierung und damit verbundene immer wiederkehrende Fragestellungen, bietet jedoch nützliche Ansätze, die es für eine mögliche Anwendung im eigenen Umfeld zu untersuchen gilt.

3 Über integration-factory

Die integration-factory ist ein stark wachsendes Unternehmen mit dynamischen Berater-Teams, das sich durch langjährige Projekterfahrung, Flexibilität im Denken und einem Höchstmaß an Motivation auszeichnet.

Seit 2004 gehört integration-factory zu den gefragten Experten für die Integration von Unternehmensdaten. Mit derzeit 48 Mitarbeitern werden namhafte Kunden aus der Finanz- und Automobil-Industrie seit Jahren in vertrauensvoller Partnerschaft erfolgreich beraten und betreut. Zu unseren Kunden gehören u.a. die EEX AG, Deutsche Börse AG, Daimler AG, Deutsche Bank AG und DZ Bank AG.

Der Fokus unserer Arbeit liegt auf Business Information Integration, also der Integration von internen und externen Unternehmensdaten, der Transformation zu nachhaltig wertvollen Informationen und deren Verteilung auch über System- und Unternehmensgrenzen hinweg. Darüber hinaus entwickeln wir umfassende Business Applikationen, die die komplexen fachlichen Problemstellungen unserer Kunden lösen und entscheidende Mehrwerte im analytischen als auch im operativen Bereich erzeugen. Wir sind verantwortlich für Geschäftsprozessanalysen, Erstellung von IT-Konzepten, IT- und Unternehmensberatung sowie Projektmanagement. Ziel ist immer die perfekte technische Umsetzung der entworfenen Datenintegrationslösungen und Business Applikationen. Unser wichtigster Differentiator ist dabei das belegbar überragende Verhältnis zwischen Kosten und Ergebnis unserer Arbeit, sowohl in quantitativer als auch qualitativer Hinsicht. Dieser Vorteil für unsere Kunden basiert auf erprobter Vorgehensweise, Methoden und Pattern mit automatisierter Umsetzung sowie einer äußerst vitalen Know-How-DNA.

In unseren Projekten lösen wir die Aufgaben in einem interdisziplinären Kontext und setzen State-of-the-Art-Technologien ein. Damit sind wir immer am Puls der Zeit bei neuen Technologien und Trends der Branche, wie aktuell Big Data und Realtime Data Warehousing.

Die internationalen Auszeichnungen TDWI-Best-Practice- und Leadership Award als auch Informatica Partner of the Year Central Europe belegen unsere erfolgreiche Arbeit.

integration-factory GmbH & Co. KG

Windmühlstraße 2

60329 Frankfurt am Main

Telefon: +49 69 25669269-0

Web: www.integration-factory.deE-Mail: info@integration-factory.de**Daniel Feidieker**

Partner

Michael Schliebitz

Manager