

Virtual Enterprise Data Warehouse oder: Warum Sie wirklich eine PSA (Persistent Staging Area) für Ihr DWH wollen

Starken Regen gibt es in Hamburg nicht wirklich häufig – zumindest gefühlt. Doch ausgerechnet in der Stunde vor dem 33. TDWI Roundtable in Hamburger hat es stark geregnet. Um so erfreulicher ist es dann, dass knapp 70 Teilnehmer es sich nicht nehmen lassen wollten, einen hervorragenden Vortrag zum Thema „Virtual Enterprise Data Warehouse“ zu besuchen. Um es vorweg zu nehmen: sie wurden nicht enttäuscht.

Vortragender des Abends war der bekannte TDWI-Roundtable Vorsitzende aus Hannover, Christian Hädrich. Christian Hädrich beschäftigt sich schon seit vielen Jahren mit Datenmodellierung und ist einer der aktiven Data Vault Protagonisten im deutschsprachigen Raum. Er ist Mitbegründer der deutschsprachigen Data Vault User Group und unterstützt auch entsprechende Veranstaltungen zum Thema Datenmodellierung.

Neben dem Regen (oder evtl. auch aufgrund des Regens) war die Verkehrslage in der Hamburger Innenstadt angespannt, weshalb man kurzfristig entschieden hat 10 Minuten später anzufangen und plante, die Vorstellung des Vereins etwas zu kürzen. Aber schon bei der Frage, wer denn alles zum ersten Mal dabei wäre, wurde klar, dass eine Kürzung der Vereinsvorstellung suboptimal wäre, da ca. 50% der Teilnehmer Erstbesucher waren. Evtl. lag es daran, dass zum ersten Mal auch MeetUp als Medium zur Bewerbung genutzt

wurde. Zu vermuten ist aber, dass es mehr am Thema lag, was später auch noch deutlich wurde.

Der Vortrag wurde mit der Feststellung begonnen, dass eine Modellierung immer nur ein Abbild der Realität darstellt und im Allgemeinen die Sicht des Modellierers wiedergibt, was anhand von Beispielen untermauert wurde. Als nächstes wurde festgehalten, dass man die gleichen Informationen unterschiedlich darstellen kann – gegebenenfalls auch muss, in Abhängigkeit der Nutzung dieser Daten. Als Beispiel hierzu dienten die Kartendarstellungen unterschiedlicher Anbieter zum Raum Hamburg oder auch die Netzpläne des HVVs. Schön an dieser Stelle ist, wie der Redner es schaffte, sein Publikum mit einfachen, aber für jeden Anwesenden deutlich zu verstehenden Beispielen, alle auf einen Stand zu bringen. Durch die Nutzung von Hamburg als Datenbasis ist dies sehr gut gelungen.

Als nächstes wurden die Strukturen des SPOT (Single Point of Truth) aufgezeigt. Auch hier wurde wieder anhand von Beispielen erläutert, wie sich Vorgänge des realen Lebens in ein Datenmodell transformieren lassen. Dem Publikum wurde verdeutlicht, dass transformierte Daten nicht immer leicht in den Ausgangszustand transformiert werden können, wenn die Transformationen nicht bekannt sind. Als Ergebnis ist festgehalten worden, dass es mehr als sinnvoll ist, die Ursprungsdaten immer zu Speichern und zwar in den PSA (Persistent Staging Area). Auf diesen aufsetzend erfolgen

dann der Integration Layer und der Interpretation Layer. Bei diesen Layern sollte man sehr behutsam mit der Persistierung arbeiten. Zum einen geht die Flexibilität verloren, die man vor allem in der Entwicklungsphase benötigt und zum anderen sind zusätzliche Speicherungen immer mit Arbeit verbunden. Solange es möglich ist, sollte man die Layer virtuell halten. Wenn man sich dann noch an die besprochenen „Fundamental ETL principles“ hält, kann die Struktur des DWH hoch automatisiert erfolgen. In dieser Phase des Vortrags wurde sehr viel diskutiert von „Muss man das wirklich virtuell halten?“ über „Wieviel Automatisierung ist denn möglich?“ bis hin zu „Wie ist das mit der Historisierung und dem Datenschutz?“.

Am Ende wurden alle Fragen in der Gruppe diskutiert und abschließend beantwortet.

Für die Teilnehmer, die nach dem Vortrag gern mal selber ausprobieren wollen, wie sich das gezeigte umsetzen lässt, hatte Christian Hädrich dann noch einige Hinweise zu Tools, die man frei im Internet downloaden kann. Diese Tools basieren auf den Ergebnissen von Roelant Vos, ein weiterer Protagonist zum Thema Data Vault und automatisierten DWH-Beladen.

Trotz der deutlichen Überziehung des Vortrags, was den vielen Diskussionen während des Vortrags geschuldet war, wurde wie immer im Nachgang bei Brötchen und kalten Getränken weiter diskutiert und wurden Erfahrungen ausgetauscht.

Das TDWI Roundtable Team Hamburg bedankt sich an dieser Stelle bei Christian Hädrich für seinen von allen Seiten gelobten Vortrag. Der Vortrag war sowohl inhaltlich sehr gut als auch von einem sehr lockeren Vortragsstil begleitet, der für die späte Stunde sehr erheiternd war.

Das TDWI Roundtable Team bedankt sich auch bei den Teilnehmern für den regen Austausch und die Hilfe beim Abräumen. Es zeigte deutlich, wie der Rahmen und die Möglichkeit zur Diskussion untereinander genutzt werden und dass dies weiterhin ein fester Bestandteil des erfolgreichen TDWI Roundtable Hamburg bleiben wird.

Die Vorsitzenden freuen sich auf den nächsten TDWI Roundtable am 07. November, wenn es um das Thema „Ethik bei der Datennutzung“ geht.

Georg Franzke