

3. Übungsblatt

Besprechung: 24.11 (Gruppe A), 1.12 (Gruppe B)

Aufgabe 1

a) Entwerfen Sie für das Star Schema in Abbildung 1, wie es für Übungsblatt 2 entworfen wurde, einen ETL Prozess, der Daten aus der AWC OLTP Datenbank in dieses Schema lädt. In diesem Aufgabenteil beschränken Sie sich dabei auf ein initiales Laden der Daten. Achten Sie dabei besonders darauf, dass der Prozess folgende Punkte sicherstellt:

- eine korrekte Berechnung der Kennzahlen der Faktentabelle,
- eine lückenloses Auffüllen der Dimensionszeilen insbesondere bei hierarchischen Attributen,
- die referentielle Integrität der Faktentabelle (Stichwort: Fremdschlüssel Beziehungen).

Beschreiben Sie dabei kurz, was die einzelnen Komponenten in ihrem ETL Prozess tun. Dabei sollten Sie nur den ETL Subtask, also die Datenextraktion, -transformation oder das Laden, beschreiben, Implementierungsdetails aber weglassen.

Für den Entwurf können Sie folgende Hilfen verwenden:

- Für Ihre Komponenten können Sie die Komponenten aus SQL Server Integration Services in Abbildung 2 als Vorlage nehmen, wobei Sie auch hier wieder den ETL Subtask beschreiben sollten, nicht aber auf die Implementierungsdetails eingehen brauchen.
- In [1, Kapitel 7] finden Sie Entwurfsmuster für zentrale Aufgaben im ETL Prozess, wobei zur Erfüllung der drei oben genannten Punkte die folgenden Entwurfsmuster wichtig sind:
 - Subsystem 3 “Extract System”
 - Subsystem 4 “Data Cleaning System”
 - Subsystem 10 “Surrogate Key Generator”
 - Subsystem 14 “Surrogate Key Pipeline”

b) An welchen Stellen müssen Sie Ihren Entwurf erweitern, wenn Daten nun inkrementell in das Star Schema geladen werden sollen? Skizzieren Sie die Erweiterung beispielhaft für die Faktentabelle und eine Dimensionstabelle. In der Skizze können Sie Komponenten für komplexere

ETL Aufgaben verwenden, wie die “langsam veränderliche Dimension” SSIS Komponente in Abbildung 2. Auch hier sollten Sie wieder die ETL Aufgabe der Komponente beschreiben und die Implementierungsdetails weglassen.

Für diesen Aufgabenteil können Sie weitere Entwurfsmuster aus [1, Kapitel 7] zur Hilfe nehmen, insbesondere die folgenden:

- Subsystem 2 “Change Data Capture System”
- Subsystem 9 “Slowly Changing Dimension Manager”
- Subsystem 13 “Fact Table Builders”

Aufgabe 2

Entwerfen und implementieren Sie das Laden der Dimensionstabelle Produkt und der Faktentabelle “Verkauf Einzelhandel” mit Hilfe der SQL Server Integration Services (SSIS). Die Faktentabelle können Sie auf die Kennzahl “Anzahl” und den nötigen Fremdschlüssel beschränken.

- Im Video2Brain Kurs “SQL Server Integration Services – Grundlagen” im Kapitel “Grundlagen” und “Data Flow” finden Sie einen Überblick, wie Projekte für SSIS aufgebaut sind und SSIS Komponenten verwendet werden können.
- Als Hilfe dazu können Sie wieder Entwurfsmuster aus [1, Kapitel 7] benutzen.

Literatur

- [1] J. Mundy, W. Thornthwaite, R. Kimball: The Microsoft Data Warehouse Toolkit (2nd edition), Wiley 2011.

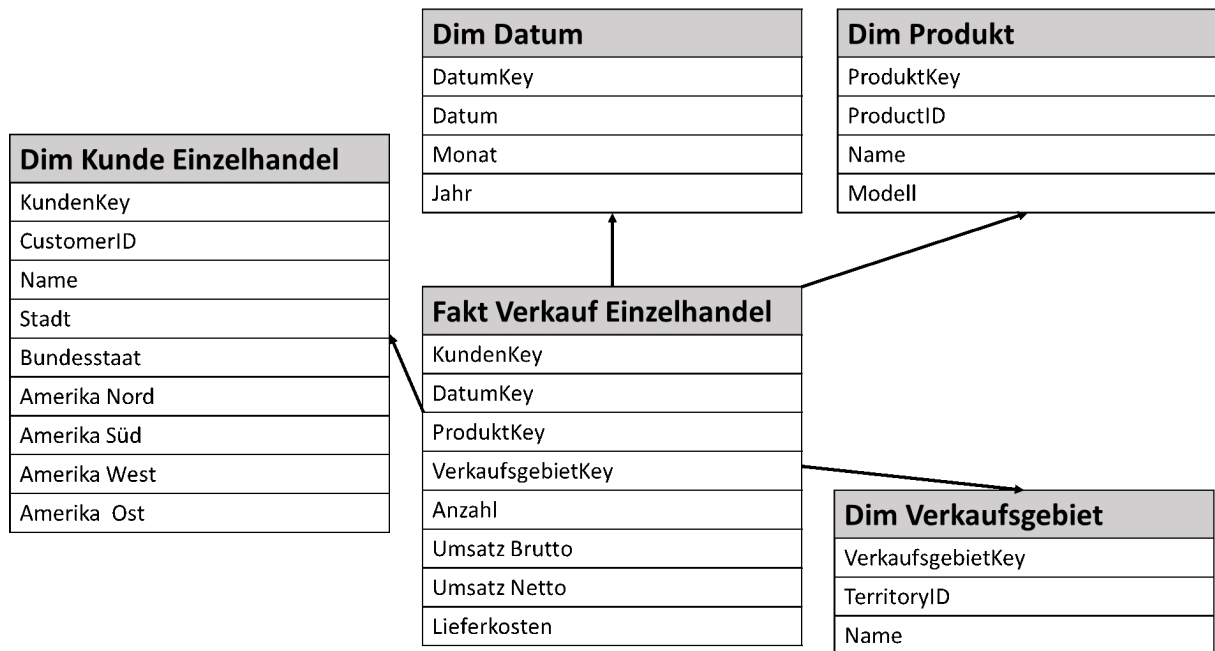


Abbildung 1: Star Schema für AWC

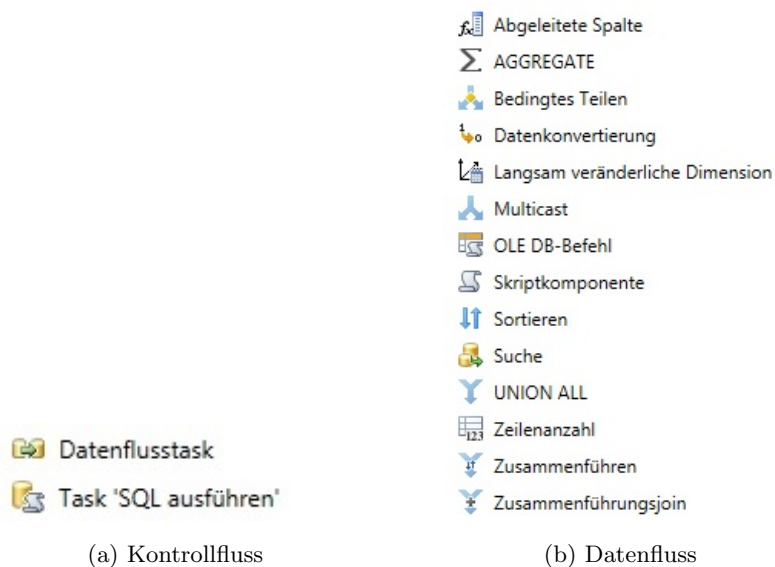


Abbildung 2: Auswahl an SQL Server Integration Services Komponenten