WI2 Aufgabe 6

Version 1

Gruppe 2

Daria Zhdanova

Martina Radic

Textuelle Erläuterungen

**ETL- Befüllen des Starschemas für den Data Marts**

Um unser Starschema aufzufüllen, haben wir mit Pentaho Data Integration eine entsprechende „ETL-Ladephase“- Job erstellt. Sie umfasst die folgenden Schritte:

1. DeltaLoad Kunden

Für „Table input“ wird mithilfe von SQL-Query auf die Tabelle DB\_SCH.WI2\_ERP10V2\_MDCUSTOMER zugegriffen.

Für die anschließende Implementierung von Slowly Changing Dimensions Typ 2 verwenden wir die Transformation "Dimension lookup/update", um die Dimensionstabelle den Kunden zu füllen. Somit fügen wir einen Surrogatschlüssel, ein Versionsattribut und die Attribute ValidFrom-ValidTo hinzu. Auf diese Weise können wir Änderungen und neue Einträge in der Produktdimension verfolgen.

1. DeltaLoad Produkte

Für „Table input“ wird hier mithilfe von SQL-Query auf die gejointe Tabellen DB\_SCH.WI2\_ERP10V2\_MDPRODUCTS und DB\_SCH.WI2\_ERP10V2\_MDPRODUCTCATEGORIES zugegriffen.

Da hier keine Historie nötig ist, verwenden wir die Transformation "Insert/update", um die Dimensionstabelle des Produkts zu füllen. Werden Änderungen an den Daten vorgenommen, wird immer der aktuelle Stand angezeigt.

1. Load ExchangeRates

Zunächst wird ein URL-Link auf die JSON-Datei als Datenquelle gespeichert. Mit der Konvertierung "JSON-Input" werden die Daten aus der JSON-Datei extrahiert und mit dem "Table Output Star" in den beiden Spalten "RateDates" und "ExchangeRates" im entsprechenden Format in der neuen Tabelle "WI2\_WOW167\_EXCHANGERATES" gespeichert.

1. DeltaLoad Faktentabelle

Als Input wurden die OLTP-Tabellen gejoint mit zusätzlicher ExchangeRates-Tabelle verwendet. Um immer aktuelle Datensätze zu haben, wird zusätzlich zu den Messschlüsseln ein Surrogatschlüssel für die Customerdimension-Tabelle verwendet.

Auch hier werden die Measures-Spalten mit der Funktion DISTINCT für NoOfSalesOrders entsprechend berechnet und dann mit den Schlüsseln Customer, Product und Orderdate für das Cube-Format gruppiert.

Um dieses Laden der Daten restartfähig zu machen, wird in der SQL-Query ein WHERE NOT EXISTS-Statement definiert, um zu prüfen, ob die Faktentabelle bereits einzufügende Datensätze enthält.

All dies wird mittels Feldmapping in die Ausgabetabelle "WI2\_WOW167\_SALESFACT" übertragen.