

ETL-Industrialisierung mit dem OWB Mapping Generator

**Irina Gotlibovych
MT AG
Ratingen**

Schlüsselworte

OWB, ETL-Automatisierung, Mapping Generierung, generische Templates, deklarative Definition

Einleitung

Wie oft in unserem Leben wünschen wir uns, bestimmte wiederkehrende Routineaufgaben nicht mehr von Hand oder am liebsten gar nicht mehr selbst erledigen zu müssen? Auch in der IT-Branche ist der Prozess der Industrialisierung nicht ausgeblieben. Im Laufe der Jahre sind viele Tools entstanden, die unsere tägliche Arbeit erleichtern und kleinere Aufgaben automatisch für uns erledigen. Im Data Warehouse Bereich werden kaum noch Prozesse manuell gestartet, sowohl das Deployment als auch die Batchläufe werden verscriptet und automatisch ausgeführt.

Wie verhält es sich jedoch mit der Entwicklung von Prozessen der DWH Datenbewirtschaftung?

ETL Entwicklung in manueller Kleinarbeit

Bei der Entwicklung von ETL-Prozessen in einem Data Warehouse sind mehrere Prozessschritte aufzubauen, die einer gleichartigen Logik folgen. So werden viele Datenobjekte aus Quellsystemen auf die gleiche Weise in den Arbeitsbereich geladen. Bei der Transformation werden Daten in das einheitliche Format der Zieldatenbank überführt (z.B. Datentypkonvertierung, Datenbereinigung). Anschließend werden Daten nach dem gleichen Prinzip (z.B. Delta Load, SCD) in das Data Warehouse eingebracht.

In der Praxis bedeutet dies oft, dass logisch identische Mappings in manueller Kleinarbeit erstellt werden. In jedem dieser Mappings müssen von Hand Operatoren angelegt und verbunden werden, für jedes Attribut eines Expression Operators muss manuell der Ausdruck eingetragen werden, Eigenschaften von Operatoren und Attributen müssen immer wieder neu gesetzt werden. Bei einer späteren Änderungsanforderung muss jedes einzelne Mapping wieder angepasst und getestet werden. Dieses Vorgehen erfordert einen hohen Entwicklungsaufwand bzw. Aufwand bei späteren Änderungen, ist fehleranfällig und testintensiv.

Wie hätten wir es gerne?

Der Nachteil bei der Entwicklung von OWB Mappings ist, dass es keine Möglichkeit gibt diese ohne Bindung an konkrete Objekte (Tabellen, Spalten usw.) anzulegen. Die fachliche Logik eines Mappings ist fest mit den Umgebungsinformationen verbunden. Um die gleiche Logik nicht mehrmals neu erzeugen zu müssen, bräuchte man einen Weg, Mappings ohne Bezug zu den eigentlichen Objekten definieren zu können. Die Erzeugung der Mappings kann dann automatisch erfolgen, wobei Objektamen als Parameter dem Generierungsprozess mitgegeben werden.

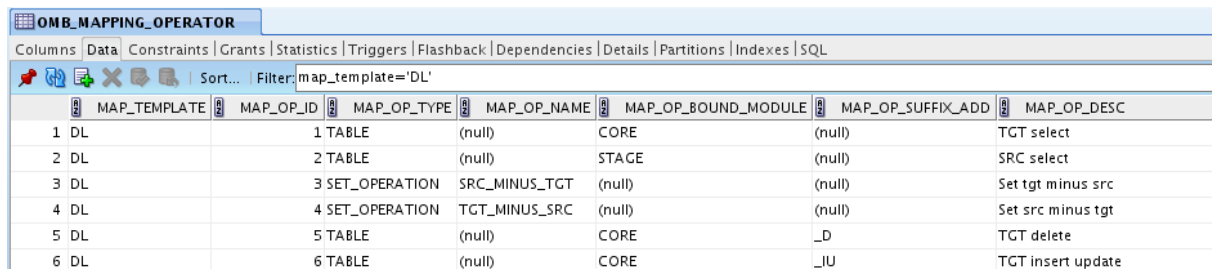
Dieser generische Ansatz wurde bei der Entwicklung des OWB Mapping Generators zu Grunde gelegt.

OWB Mapping Generator

Der OWB Mapping Generator ist ein mit Oracle- und OWB Mitteln entwickeltes Framework, welches es Ihnen ermöglicht OWB Mappings auf Basis von Templates automatisch zu generieren. Diese Templates bilden die fachliche Logik Ihrer ETL-Prozesse ab und werden in der Datenbank deklarativ anhand von Metadaten definiert. Die Definition der Templates erfolgt generisch, d.h. sie legen die fachliche Logik der Mappings fest, sind jedoch unabhängig von konkreten Objekten. Da die Generierung der Mappings später mithilfe eines TCL (OMB Plus) Skriptes erfolgt, sind für die Definition von Templates minimale OMB Plus Kenntnisse erforderlich. Um komplexe Logiken abbilden zu können werden tiefere TCL Kenntnisse benötigt.

Der OWB Mapping Generator stellt in der Datenbank einen Satz von Definitionstabellen bereit, in denen die Templates deklarativ beschrieben werden. In diesen Tabellen finden Sie keine Tabellen- oder Attributnamen, die konkreten Objekte werden erst während der Generierung automatisch an die Templates gebunden.

In den folgenden Abschnitten werden die Definitionstabellen des OWB Mapping Generators anhand von Beispiel-Templates (Delta Load Template) erklärt. Die erste Tabelle ist OMB_MAPPING_OPERATOR, hier definieren Sie Operatoren für Ihre Mappings. Die Begrifflichkeiten im OWB Mapping Generator sind die gleichen wie im Oracle Warehouse Builder – man findet sich demnach sogleich zurecht. Als Operatoren können also alle Objekte verwendet werden, die in der Komponenten-Palette zur Verfügung stehen.



	MAP_TEMPLATE	MAP_OP_ID	MAP_OP_TYPE	MAP_OP_NAME	MAP_OP_BOUND_MODULE	MAP_OP_SUFFIX_ADD	MAP_OP_DESC
1	DL	1	TABLE	(null)	CORE	(null)	TGT select
2	DL	2	TABLE	(null)	STAGE	(null)	SRC select
3	DL	3	SET_OPERATION	SRC_MINUS_TGT	(null)	(null)	Set tgt minus src
4	DL	4	SET_OPERATION	TGT_MINUS_SRC	(null)	(null)	Set src minus tgt
5	DL	5	TABLE	(null)	CORE	_D	TGT delete
6	DL	6	TABLE	(null)	CORE	_JU	TGT insert update

Abb. 1: Definitionstabelle für Operatoren

In der OMB_MAPPING_OPERATOR_PROPERTY Tabelle können Sie Eigenschaften für Mappings, Operatoren, Gruppen und Attribute definieren. Analog zum Oracle Warehouse Builder müssen nur die Eigenschaften beschrieben werden, die nicht automatisch erzeugt werden können: wenn Sie z.B. im Oracle Warehouse Builder einen Tabellen-Operator mit einem Expression-Operator verbinden, werden die Attribute der Eingabegruppe des Expression-Operators mit den richtigen Datentypen automatisch generiert. Der OWB Mapping Generator arbeitet auf die gleiche Weise. Die PROPERTY-Tabelle kann je nach Anforderung oder Komplexität der umzusetzenden Logik sowohl statische als auch dynamische Werte enthalten. Eine Eigenschaft kann mithilfe vordefinierter dynamischer Parameter definiert werden. Erfordert die fachliche Logik eine umfassendere Berechnung der Werte, z.B. abhängig vom Primärschlüssel der Tabelle oder von Datentypen der Attribute, haben Sie im OWB Mapping Generator die Möglichkeit eine benutzerdefinierte Funktion anzulegen, die dann in der PROPERTY-Tabelle verwendet werden kann.

OMB_MAPPING_OPERATOR_PROPERTY											
Columns	Data	Constraints	Grants	Statistics	Triggers	Flashback	Dependencies	Details	Partitions	Indexes	SQL
Filter: mopr_template='DL'											
	MOPR_TEMPLATE	MOPR_OP_ID	MOPR_ATTR_NAME	MOPR_OP_GROUP_TYPE	MOPR_OP_GROUP_ID	MOPR_PROP_NAME	MOPR_PROP_VALUE				
1	DL	3 (null)	(null)	(null)	(null)	SET_OPERATION	MINUS				
2	DL	4 (null)	(null)	(null)	(null)	SET_OPERATION	MINUS				
3	DL	5 (null)	(null)	(null)	(null)	LOADING_TYPE	DELETE				
4	DL	6 (null)	(null)	(null)	(null)	LOADING_TYPE	INSERT/UPDATE				

Abb.2: Definitionstabelle für Eigenschaften

Die Tabelle OMB_MAPPING_CONNECTION enthält Verbindungen zwischen den Mapping Operatoren. Die Definition erfolgt nach dem gleichen Prinzip wie im Oracle Warehouse Builder direkt. Die Reihenfolge, in der die Übergänge später erzeugt werden (wichtig für die automatische Generierung von Attributen) wird bestimmt durch die Spalte CONN_ID.

OMB_MAPPING_CONNECTION											
Columns	Data	Constraints	Grants	Statistics	Triggers	Flashback	Dependencies	Details	Partitions	Indexes	SQL
Filter: conn_template='DL'											
	CONN_TEMPLATE	CONN_ID	CONN_OP_FROM_ID	CONN_OP_TO_ID	CONN_OP_FROM_GROUP_ID	CONN_OP_TO_GROUP_ID	CONN_TYPE	CONN_DESC			
1	DL	1	1	3	1	2	COPY ALL	tgt to tgt_minus_src			
2	DL	2	1	4	1	1	COPY ALL	tgt to src_minus_tgt			
3	DL	3	2	3	1	1	COPY ALL	src to tgt_minus_src			
4	DL	4	2	4	1	2	COPY ALL	src to src_minus_tgt			
5	DL	5	3	5	1	1	BY NAME	tgt_minus_src to tgt_d			
6	DL	6	4	6	1	1	BY NAME	src_minus_tgt to tgt_ju			

Abb.3: Definitionstabelle für Übergänge

Da die Definition von Templates generisch erfolgt, brauchen wir nun einen Weg diese mit den erforderlichen OWB Objekten (Module, Tabellen usw.) zu verbinden. Um die Generierung von einzelnen Mappings entsprechend Ihrer Anforderungen zu ermöglichen, können Sie im OWB Mapping Generator Namenskonventionen und Umgebungsinformationen ablegen. Dabei spielt der Begriff „Tabellensamm“ (table radical) eine zentrale Rolle. Damit ist der gemeinsame Teil der Tabellennamen über alle im Mapping verwendeten Module hinweg gemeint. Der Tabellensamm wird bei der Generierung von Mappings verwendet, um zusammengehörende Objekte in einem Mapping zu verbinden. Die Funktionsweise des Frameworks basiert auf der Annahme, dass alle verwendeten Datenbankobjekte einer allgemeinen Namenskonvention folgen. In den abgebildeten Tabellen beschreiben Sie Namenskonventionen der Datenbankobjekte (OWB Module) und legen die Namenskonvention für die zu erzeugenden Mappings fest. Dabei können die im OWB Mapping Generator zur Verfügung stehenden dynamischen Parameter verwendet werden.

OMB_STATIC_DB_NAMING										
Columns	Data	Constraints	Grants	Statistics	Triggers	Flashback	Dependencies	Details	Partitions	Indexe
Filter:										
	OWB_MODULE_ID	OWB_MODULE	OWB_MODULE_ABBR	TABLE_NAMING						
1	SOURCE	SOURCE	SRC	\$owb_module_abbr_\$tab_radical						
2	STAGE	STAGE	STG	\$owb_module_abbr_\$tab_radical						
3	CORE	CORE	(null)	\$tab_radical						

Abb.4: Namenskonventionen der Datenbankobjekte

OMB_STATIC_MAP_NAMING		
Columns	Data	Constraints
Sort... Filter:		
1	LOAD_ALL	\$business_area_\$sub_area_\$tab_radical_\$template

Abb.5: Namenskonventionen der OWB Objekte

Damit entwickeln Sie nun nicht mehr jedes einzelnes Mapping manuell im OWB Design Center, sondern definieren ein Template für eine Mapping-Klasse. Anschließend generieren Sie die dazugehörigen Mappings unter Einbeziehung Ihrer Projektvorgaben automatisch. Für die Generierung der Mappings steht im OWB Mapping Generator ein Expert (OWB Mini-Applikation) zur Verfügung, der Sie dialoggestützt durch die einzelnen Schritte leitet. Nachdem Sie Ihre Auswahl bzgl. des zu generierenden Templates und der zu verwendeten Objekte getroffen haben, generiert der OWB Mapping Generator ein TCL (OMB Plus) Skript, mit dem die Mappings automatisch erzeugt werden. Das Ergebnis ist sofort im Oracle Warehouse Builder sichtbar und kann weiterverwendet bzw. deployed werden.

OWB Mapping Generator

Der OWB Mapping Generator wurde bereits erfolgreich in der Praxis eingesetzt. Dadurch, dass die Definition von Templates exakt der Struktur von OWB-Mappings folgt, ist eine Einarbeitung in das Framework sehr schnell möglich. Die Vorteile liegen in deutlicher Reduzierung der Entwicklungszeit sowie Vereinheitlichung und damit Qualitätsverbesserung des Codes. Mithilfe des OWB Mapping Generators kann ein Data Warehouse in kurzer Zeit neu aufgebaut sowie auf geänderte Anforderungen sehr schnell reagiert werden.

Kontaktadresse:

Irina Gotlibovych
 MT AG
 Balcke-Dürr-Allee, 9
 D-40882 Ratingen

Telefon: +49 (0) 2102 309 61-0
 Fax: +49 (0) 2102 309 61-10
 E-Mail: irina.gotlibovych@mt-ag.com
 Internet: www.mt-ag.com