

# CALUMMA ETL Tool

## Benutzerhandbuch



# Inhaltsverzeichnis

<b>Teil I Einleitung</b>	<b>3</b>
<b>Teil II Spoon Basics</b>	<b>4</b>
<b>Teil III CALUMMA Steps</b>	<b>9</b>
1 Master Data Importer.....	10
2 Data Importer.....	11
3 Enumeration Mapping.....	12
4 Hierarchie-Mapping.....	15
5 Crosslink-Importer.....	16
6 Löschen beim Import.....	16

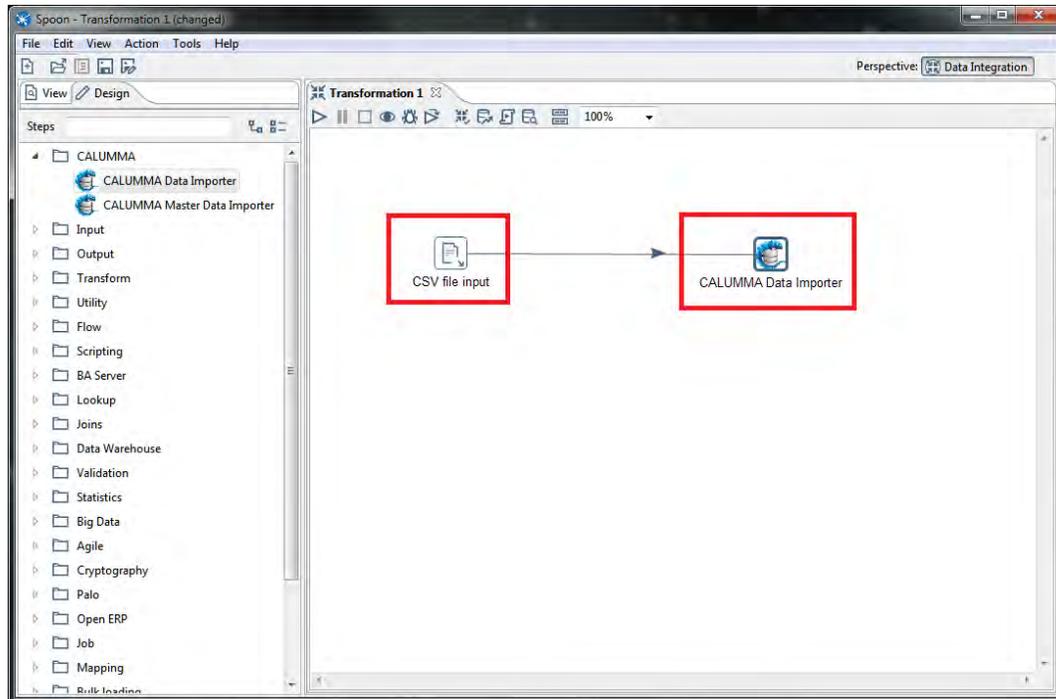
# 1 Einleitung

Das CALUMMA ETL (Extract-Transform-Load) Tool besteht aus mehreren Plug-ins für die Open Source Software *Spoon* (Kettle Data Integration) von *Pentaho* und kann zusammen mit dem CALUMMA Management Tool verwendet werden, um Daten aus einer beliebigen Datenquelle in eine CALUMMA Instanz zu kopieren. In diesem Handbuch wird die Version Spoon 7.1 verwendet.

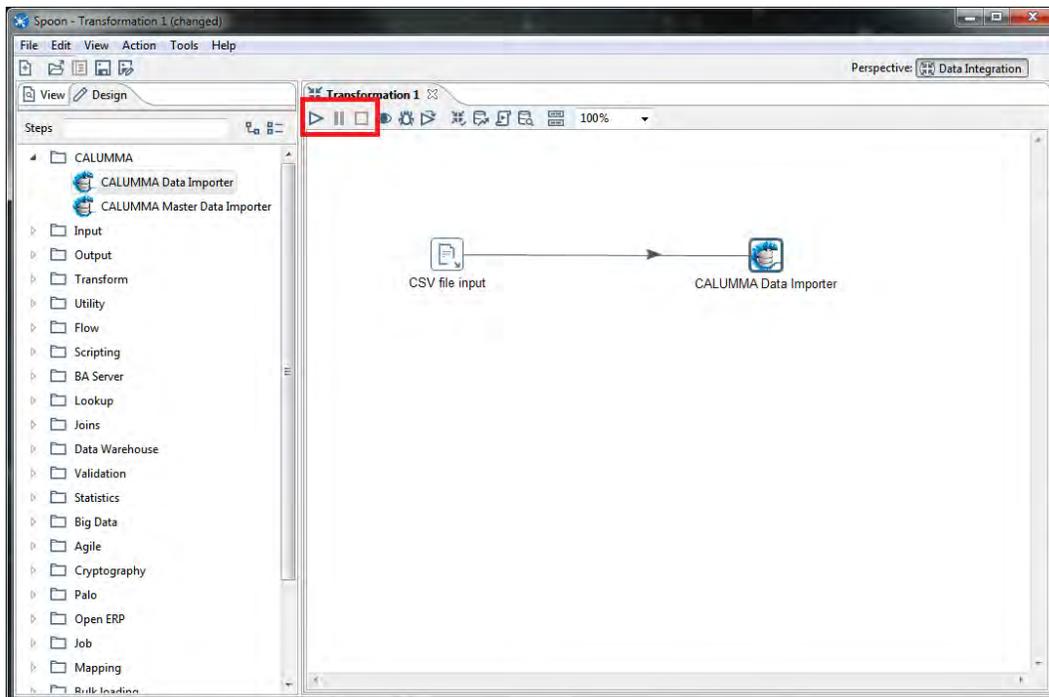
In den folgenden Kapiteln werden die grundlegenden Funktionen von Spoon erklärt, um die CALUMMA Plug-ins verwenden zu können. Allerdings bietet Spoon viele zusätzliche Funktionen, die hier nicht erwähnt werden. Mehr Informationen und eine Benutzerdokumentation zu Spoon findet man auf der Website von Pentaho unter <http://community.pentaho.com/projects/data-integration/>.

## 2 Spoon Basics

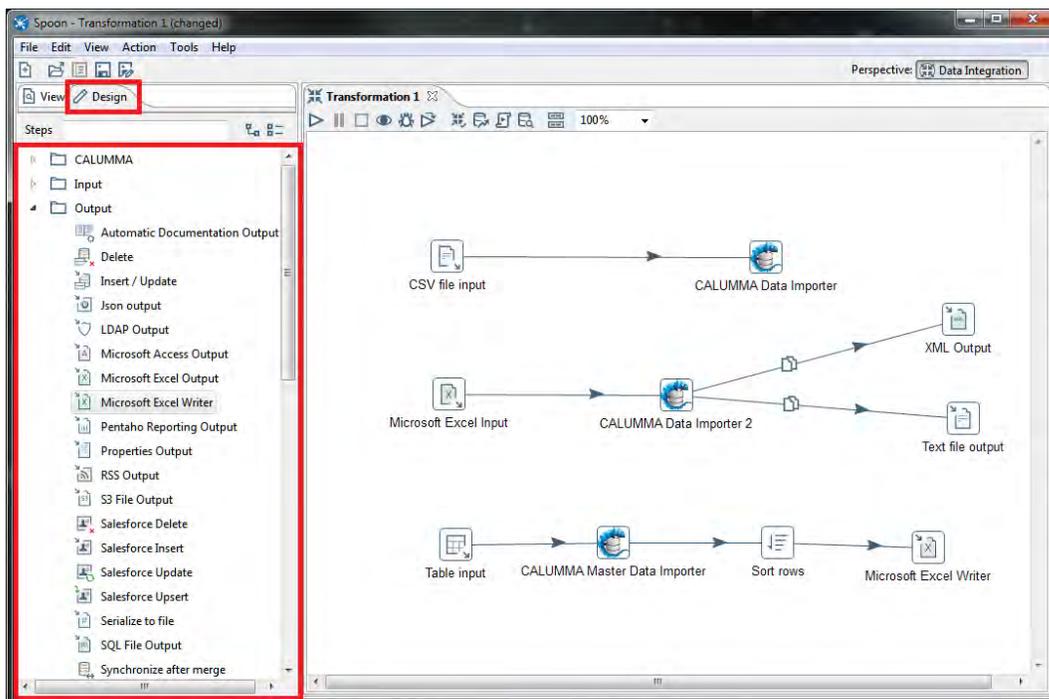
In Spoon unterscheidet man zwischen *Jobs* und *Transformationsen*, wobei ein Job mehrere Transformationen ausführen kann. Eine Transformation besteht aus einem oder mehreren Steps, welche innerhalb der Transformation bestimmte Aufgaben ausführen, zum Beispiel Daten laden, konvertieren oder speichern. In folgendem Screenshot sieht man die Transformationsansicht mit den Steps "CSV Input" und "CALUMMA Data Importer".



Links oben in der Transformationsansicht befindet sich die Symbolleiste für Transformationen. Hier kann die Transformation gestartet, pausiert und gestoppt werden.

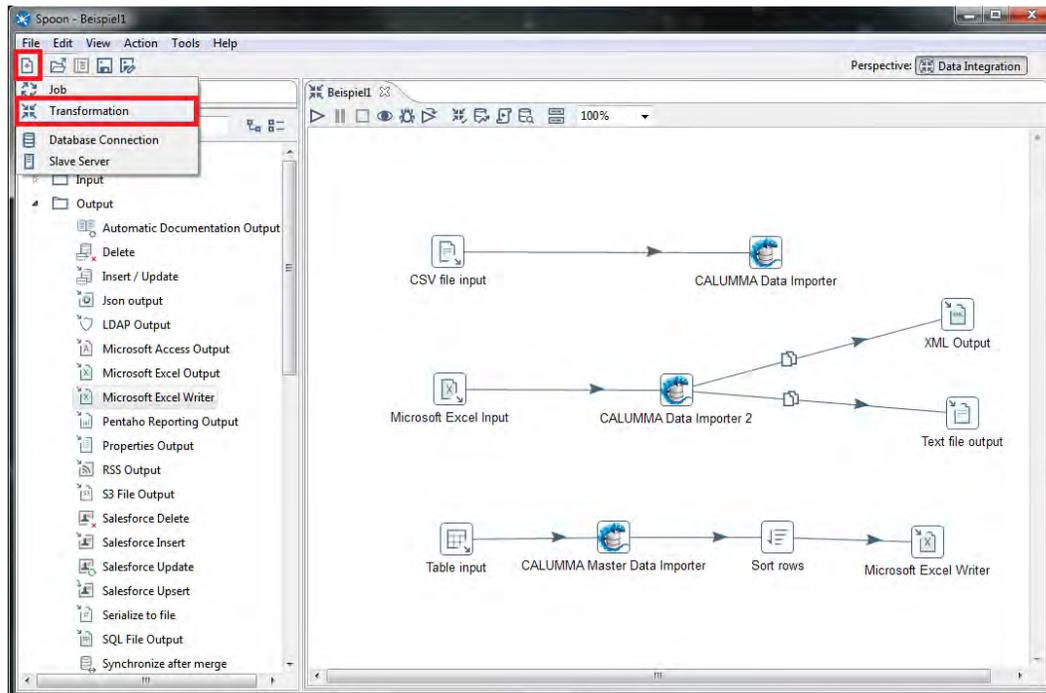


Links im "Design"-Tab befinden sich alle Steps zum Importieren, Exportieren und Transformieren von Daten. Zu jeder Transformation können beliebig Steps hinzugefügt werden.

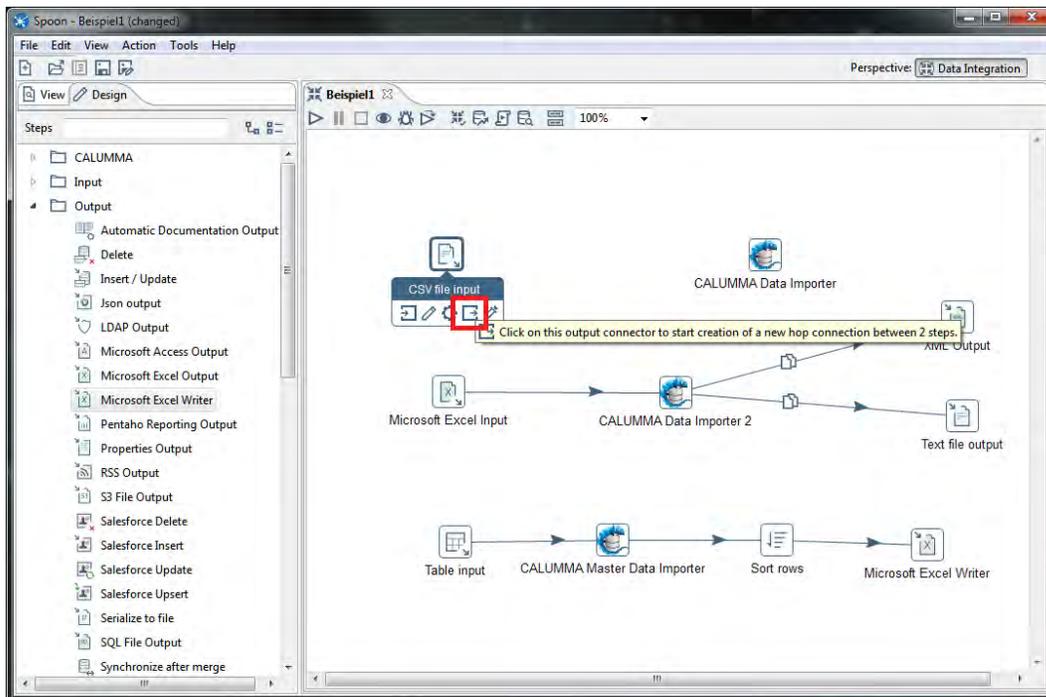


## Neue Transformationen erstellen

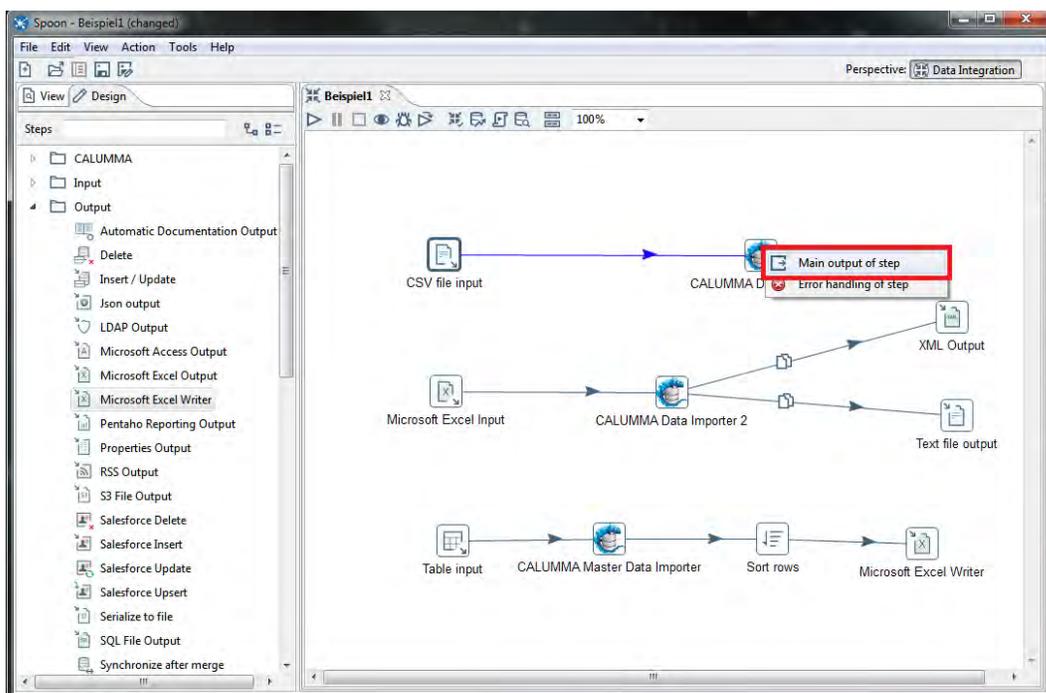
Um eine neue Transformation zu erstellen klickt man in der Menüleiste auf *File -> New -> New Transformation* oder auf das Symbol in der Symbolleiste. Eine neue leere Transformation öffnet sich danach in der Transformationsansicht.



Um Steps zu einer Transformation hinzuzufügen werden sie einfach per Drag&Drop von links vom "Design"-Tab in die Transformationsansicht gezogen. Bevor die Transformation gestartet werden kann, müssen die Steps noch miteinander verbunden werden. Dies funktioniert entweder, indem man mit der mittleren Maustaste auf den ersten Step klickt und dann mit der linken Maustaste auf den zweiten Step, oder indem man den Button "Output Connector" unter dem ersten Step verwendet und dann mit gedrückter linker Maustaste eine Verbindung zum zweiten Step herstellt. Die Verbindungen zwischen den Steps nennt man *Hops*.



Sobald man den Hop erstellt hat muss man angeben, ob in den Step die Fehlermeldungen oder der Main-Output geschrieben werden soll. Mit einem Doppelklick auf einen Step kann dieser konfiguriert werden.



## Steps löschen

Um einen Step zu löschen klickt man mit der rechten Maustaste darauf und wählt "Delete Step" oder man wählt den Step mit der linken Maustaste aus und drückt "DEL".

## CSV-Import

Folgende Dokumentation wurde mit Version Kettle 7.1\_2017\_10\_25 erstellt.

Der CSV-Importer importiert Daten aus einer CSV-Datei. Bei "Filename" muss der Pfad zur Datei angegeben werden und zusätzlich muss noch das Trennzeichen vom CSV-File angegeben werden. Danach klickt man auf "Get Fields" um die Felder zu importieren. Der Datentyp wird automatisch erkannt, jedoch sollte immer überprüft werden ob tatsächlich der richtige Datentyp ausgewählt wurde. Speziell beim Datum muss dann eventuell noch das Format angegeben werden wie: "dd.MM.yyyy" oder "dd.MM.yyyy HH:mm:ss".

CSV Input dialog box configuration:

- Step name: Kunden
- Filename: C:\develop\calumma\EclipseRCP\_CalummaV2.1\EclipseRCP\_CalummaV2.1\workspace\calumma\Additional\doc\Har
- Delimiter: ;
- Enclosure: "
- NIO buffer size: 50000
- Lazy conversion?:
- Header row present?:
- Add filename to result:
- The row number field name (optional):
- Running in parallel?:
- New line possible in fields?:
- File encoding: UTF-8

#	Name	Type	Format	Length	Precision	Currency	Decimal	Group	Trim type
1	field	String							none

Buttons: Help, OK, Get Fields, Preview, Cancel

### 3 CALUMMA Steps

Folgende CALUMMA-Steps stehen zur Verfügung und sollten in dieser Reihenfolge ausgeführt werden.

Master-Data-Importer, Data-Importer und Crosslink-Importer.

#### Konfiguration der Steps

Im "Config"-Tab werden die Daten für die Datenbankverbindung eingegeben. Zusätzlich müssen Benutzername und Passwort vom CALUMMA Management Tool Benutzer eingegeben werden.

#### Optionen

Zum Schluss können einige Optionen aktiviert oder deaktiviert werden.

Option	Funktion
Block Database on Import	Versetzt die Datenbank in den "Maintenance Mode". In diesem Modus werden alle Records gesperrt und können nicht bearbeitet werden.
Allow Mapping to internal ID	Nach dem Aktivieren dieser Funktion ist es möglich einen Wert auf die interne ID einer Entity zu mappen.
Write in Audits	Definiert, ob das ETL Tool in die Audits des Management Tools schreiben soll.

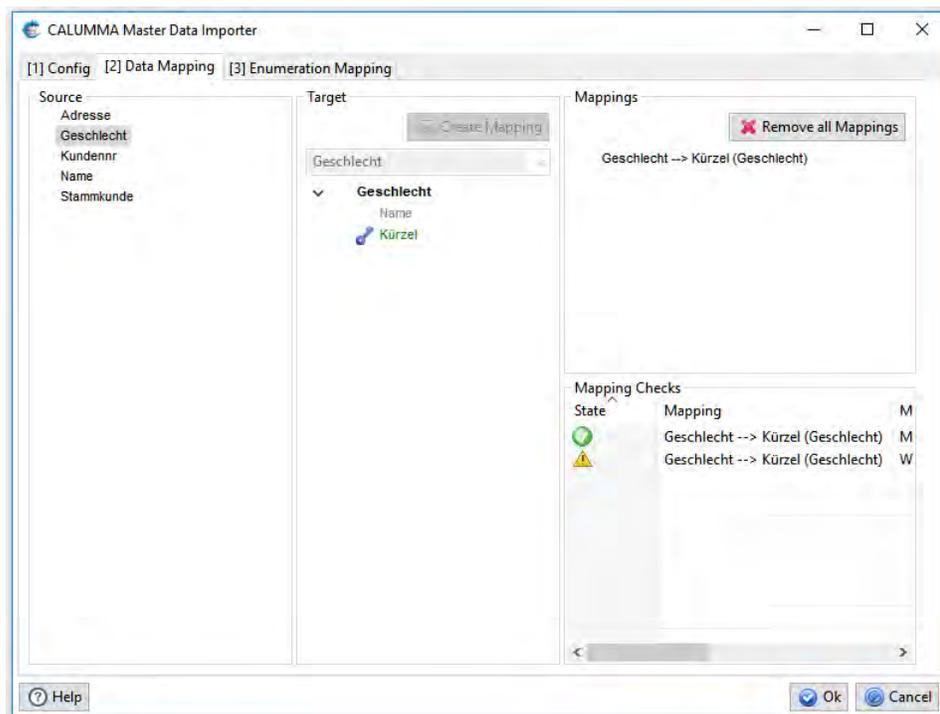
Sort alphabetically	Ändert die Sortierreihenfolge der Calumma-Attribute (alphabetisch oder die Reihenfolge die im Calumma definiert wurde)
Allow Deletes	Löschen aktivieren/deaktivieren
ChangeLog	Änderungen verfolgen aktivieren/deaktivieren

### 3.1 Master Data Importer

Der Master Data Importer kann verwendet werden um Enumerationswerte im CALUMMA Management Tool zu importieren. Um den Master Data Importer zu konfigurieren muss dieser mit Doppelklick geöffnet werden.

#### Data Mapping

Im "Data Mapping"-Tab können die Mappings erstellt werden. In der "Source"-Group werden alle Felder angezeigt, die aus dem "Input-Step" geladen werden, zum Beispiel Spaltenüberschriften von einer Excel-Datei. Zuerst wählt man in der "Source"-Group das Feld, welches die Werte für auf die zu mappenden Enumeration enthält. Danach kann von der Combobox in der "Target"-Group die Enumeration ausgewählt werden. Darunter werden die Attribute der Enumeration angezeigt. Zum Schluss muss noch ausgewählt werden, auf welches Attribut gemappt werden soll und welches von den Attributen als Primärschlüssel verwendet werden soll. Danach kann auf "Create Mapping" geklickt werden.

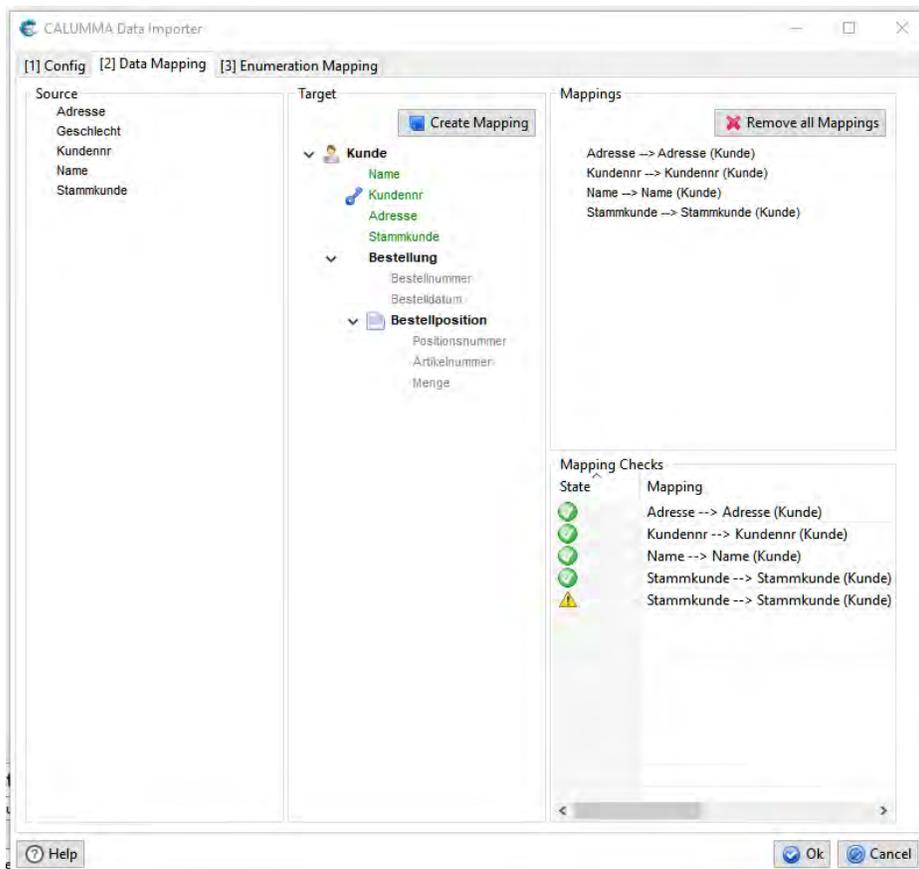


Pro Step können nur Mappings für eine Enumeration erstellt werden. Für mehrere Enumerationen müssen mehrere Steps verwendet werden.

## 3.2 Data Importer

Der Data Importer kann verwendet werden, um Daten aus einer beliebigen Datenquelle in eine CALUMMA Datenbank zu importieren.

Der folgende Screenshot zeigt den "Data Mapping"-Tab. Links werden die Felder aus der Source-Datei angezeigt, rechts der Entity-Baum aus dem CALUMMA Management Tool. Um Mappings zu erstellen klickt man auf ein Feld aus der "Source"-Liste und auf ein Attribut einer Entity, auf welche gemappt werden soll und abschließend auf "Create Mapping". Die Mappings werden ganz rechts unter "Mappings" angezeigt. Für jede Entity (außer Kompositionen) muss ein Primärschlüssel definiert werden, wenn auf ein Attribut der Entity gemappt wird. Schlüssel werden mit Doppelclick auf das Attribut erstellt. Es gibt 3 Arten von Primärschlüsseln:



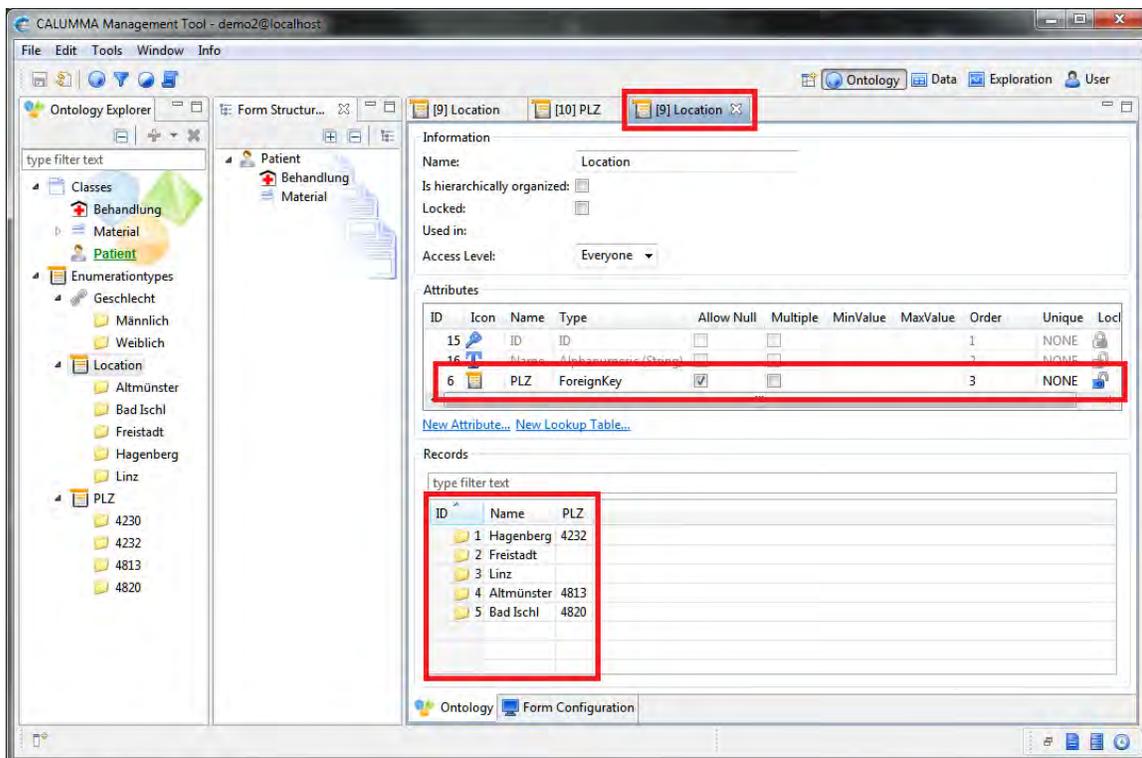
- Normaler Schlüssel (dunkelblau) -> Doppelclick
  - Create und Update von Records ist erlaubt

- Schwacher Schlüssel (hellblau) -> 2mal Doppelklick
  - nur Update von Records ist erlaubt
- Delete Schlüssel (schwarzer Totenkopf) -> 3mal Doppelklick
  - Create, Update und Delete ist erlaubt (Löschen siehe Kapitel [Löschen beim Import](#)<sup>16)</sup>)

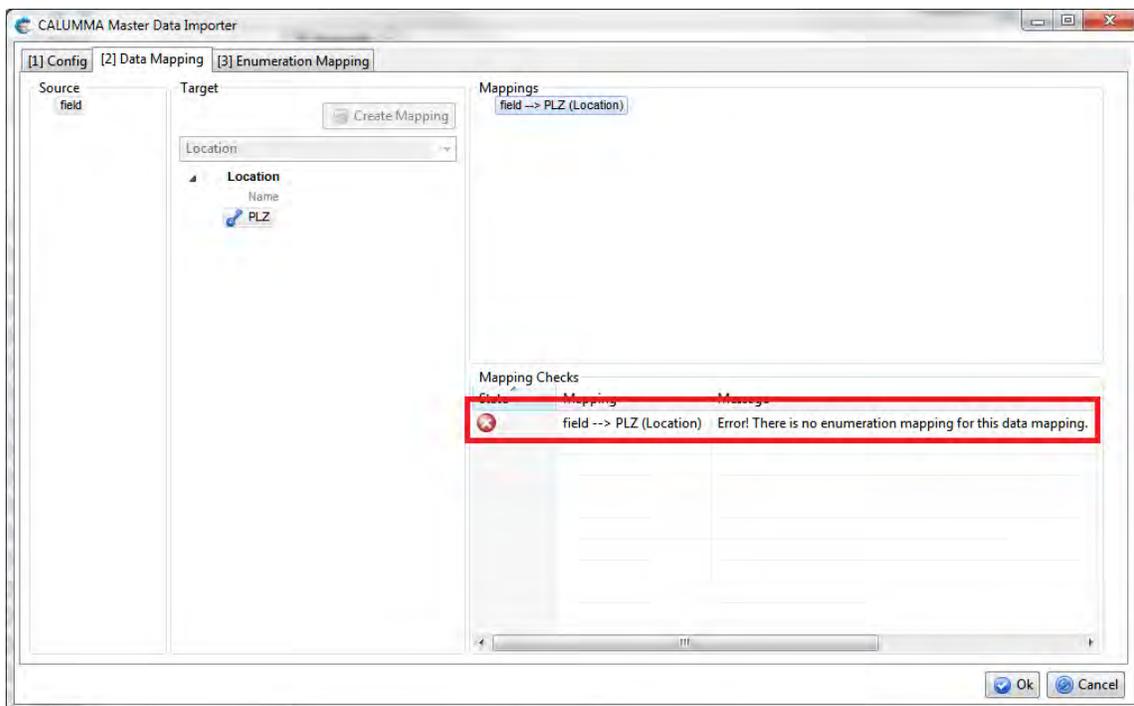
Rechts unten werden die Mapping-Checks angezeigt. Hier erscheinen Fehlermeldungen für Mappings, die nicht richtig oder noch nicht vollständig gemappt wurden, oder falls kein Primärschlüssel definiert wurde.

### 3.3 Enumeration Mapping

Folgender Screenshot vom Management Tool zeigt die Enumeration "Location". Es wurde eine weitere Enumeration "PLZ" zu "Location" hinzugefügt. Um auf diese Enumeration mappen zu können, muss ein "Enumeration Mapping" erstellt werden.



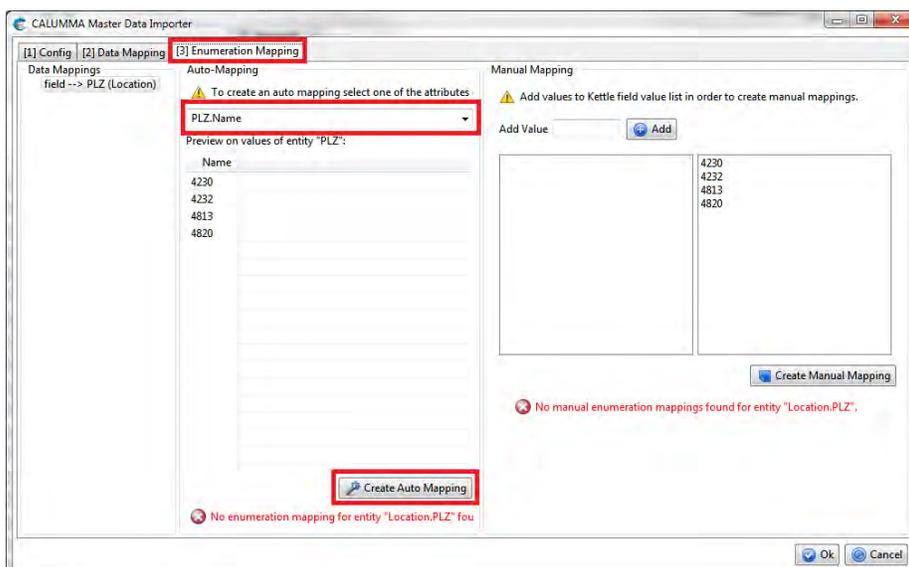
Im ETL Tool wird nachfolgendes Mapping erstellt. Da "PLZ" eine Enumeration ist erscheint in den "Mapping Checks" eine Fehlermeldung die darauf hinweist, dass ein "Enumeration Mapping" notwendig ist.



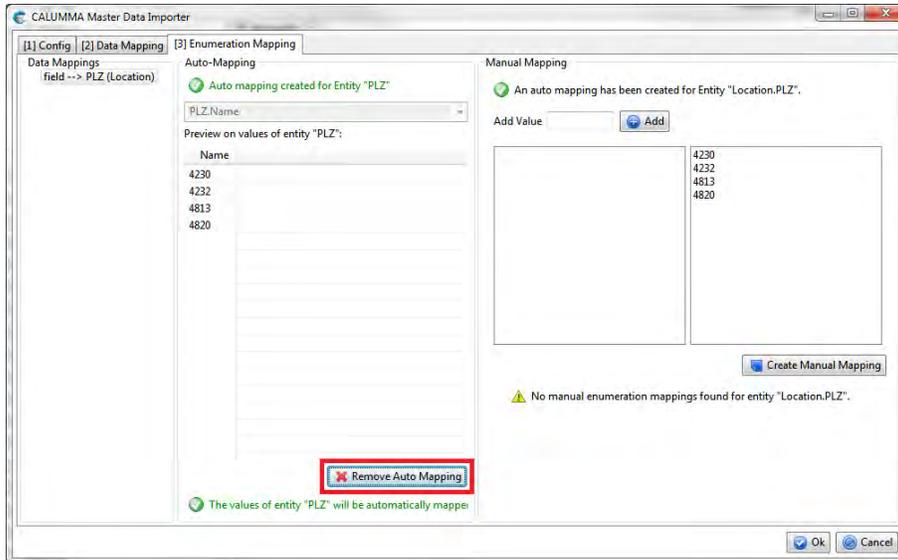
## Auto Mapping

Im "Enumeration Mapping"-Tab werden links jene Data Mappings angezeigt, für welche ein "Enumeration Mapping" notwendig ist - in diesem Beispiel ist das nur eines. Es kann ein "Auto-Mapping" oder ein "Manuelles Mapping" erstellt werden. Die Wahl ist abhängig von der Datenquelle. Wenn die Daten aus der Datenquelle in der gleichen Form sind wie im Management Tool, dann ist ein "Auto-Mapping" möglich.

Um ein "Auto-Mapping" zu erstellen wählt man aus der Combobox das Attribut aus, auf welches man Mappen möchte und klickt auf "Create Auto Mapping".

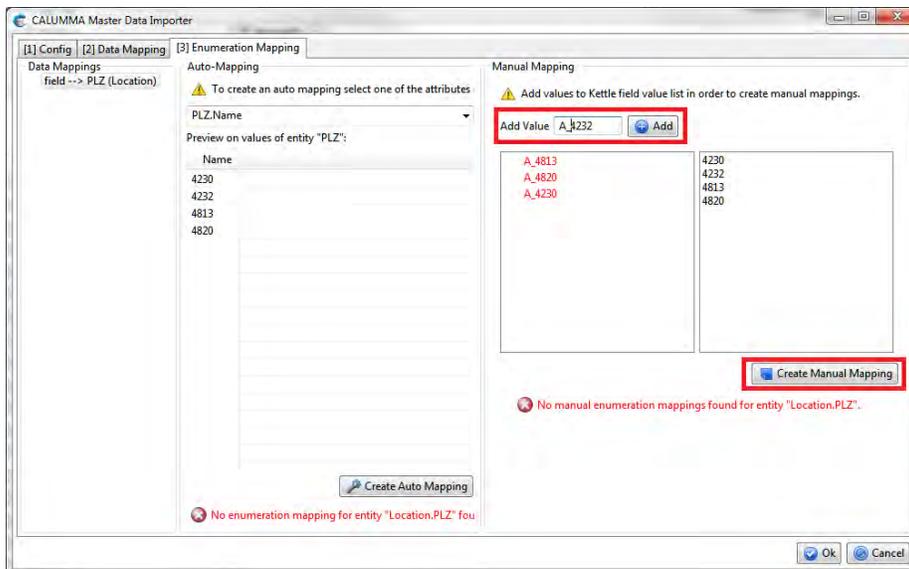


Um ein "Auto-Mapping" wieder zu löschen klickt man auf den Button "Remove Auto Mapping".



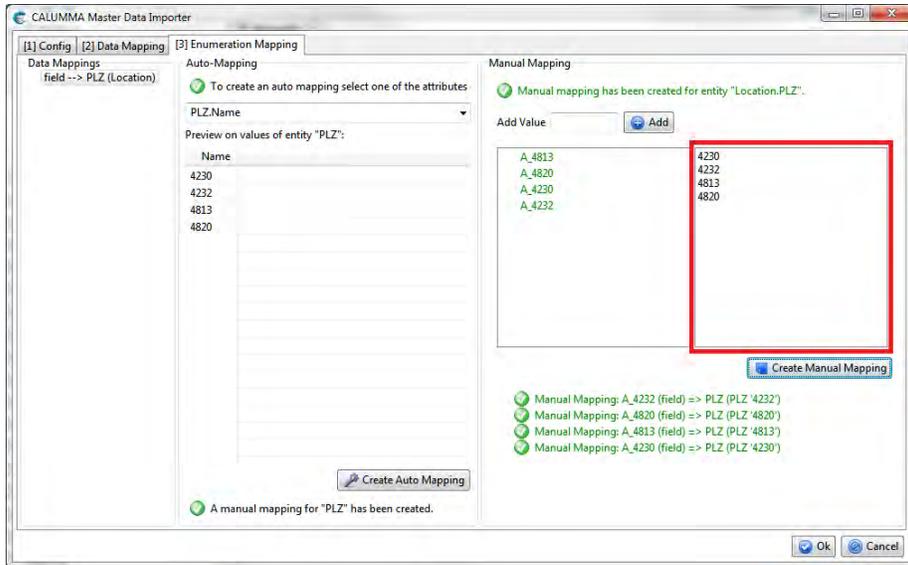
### Manual Mapping

Falls die zu mappenden Enumerationswerte nicht genau der Form entsprechen wie sie im Management Tool eingegeben wurden, können "Manual Mappings" erstellt werden. Dazu müssen die Werte hinzugefügt werden, welche gemappt werden sollen. Im nachfolgenden Screenshot werden Werte für PLZ eingefügt.



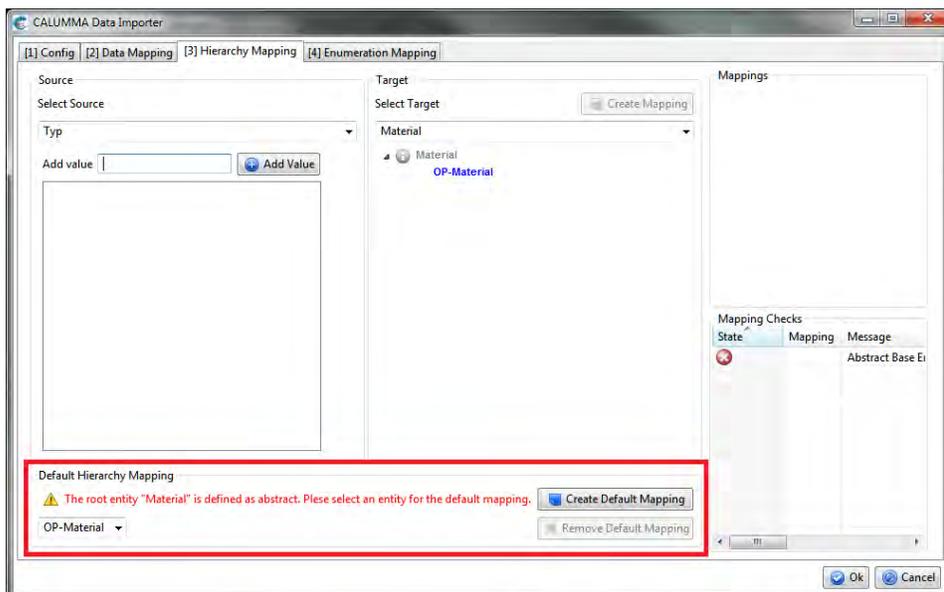
In der rechten Liste werden jene Werte angezeigt, welche im Management Tool gespeichert wurden. Um ein Mapping zu erstellen klickt man auf den erstellten Wert in der linken Liste,

dann auf den zugehörigen Wert in der rechten Liste und abschließend auf "Create Manual Mapping". Die gemappten Werte werden in grün dargestellt.



### 3.4 Hierarchie-Mapping

Wenn in der Ontologie eine Entity mit Sub-Entities definiert wurde, erscheint zusätzlich ein "Hierarchie-Mapping"-Tab. Falls die Root-Entity abstrakt ist muss ein "Default-Hierarchie-Mapping" erstellt werden. Damit wird - falls es kein "Hierarchie-Mapping" gibt - ein Record von dieser Entity erstellt. Im Beispiel unten ist die Entity "Material" abstrakt, also darf von dieser Entity kein Record erstellt werden. Aufgrund des "Default-Hierarchy-Mappings" werden automatisch alle Materialien zu OP-Material.

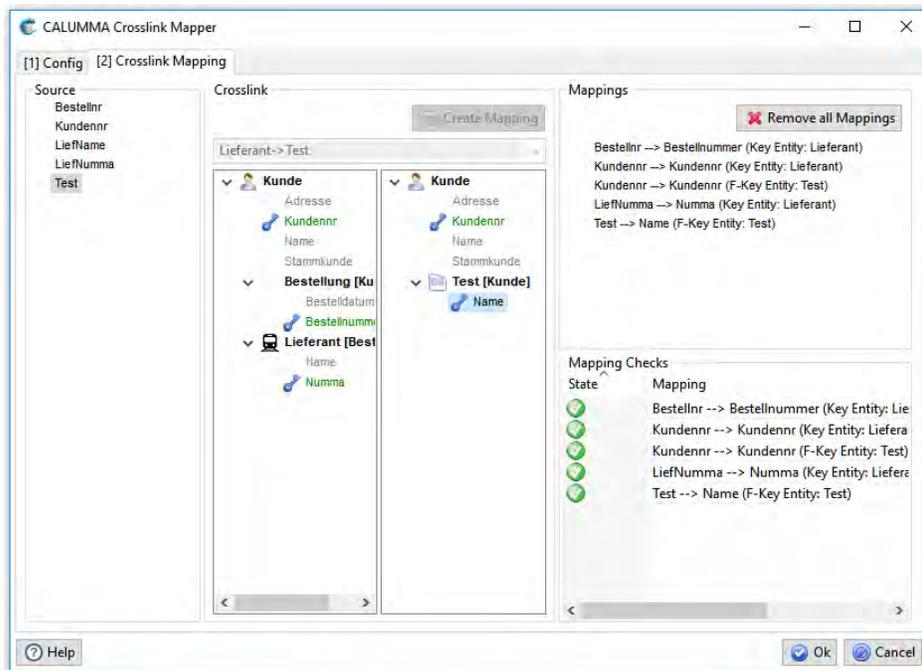


Um ein "Hierarchie-Mapping" zu erstellen wählt man unter "Select Source" zunächst das gewünschte Mapping-Feld aus der Combobox aus. Danach können zu diesem Feld *Values* hinzugefügt werden, indem man auf "Add Value" klickt. Unter "Target" wählt man die Entity aus, auf welche die Values gemappt werden. Für abstrakte Entities kann kein Mapping erstellt werden, diese werden in grau angezeigt. Um das Mapping zu erstellen muss auf die Schaltfläche "Create Mapping" geklickt werden.

### 3.5 Crosslink-Importer

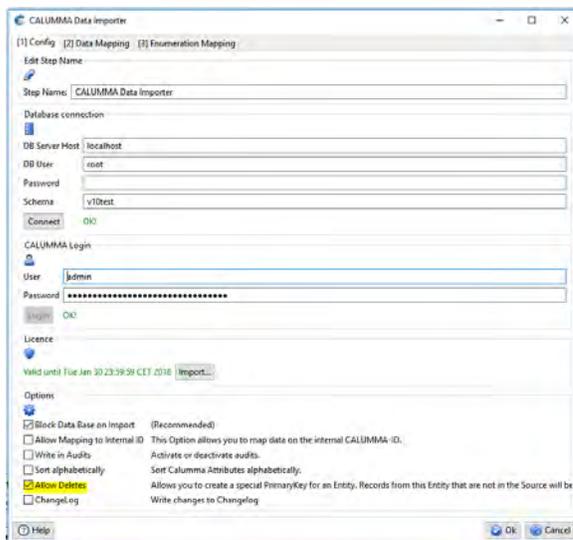
Wurde im CALUMMA Management-Tool ein Crosslink zwischen Entities erstellt, können mit dem Crosslink-Importer automatisch Records verlinkt werden.

In der Combobox wählt man den Crosslink aus, darunter werden die Entity-Attribute Trees angezeigt. Hier müssen keine Schlüsselfelder definiert werden, jedes gemappte Feld ist automatisch ein Schlüsselfeld.

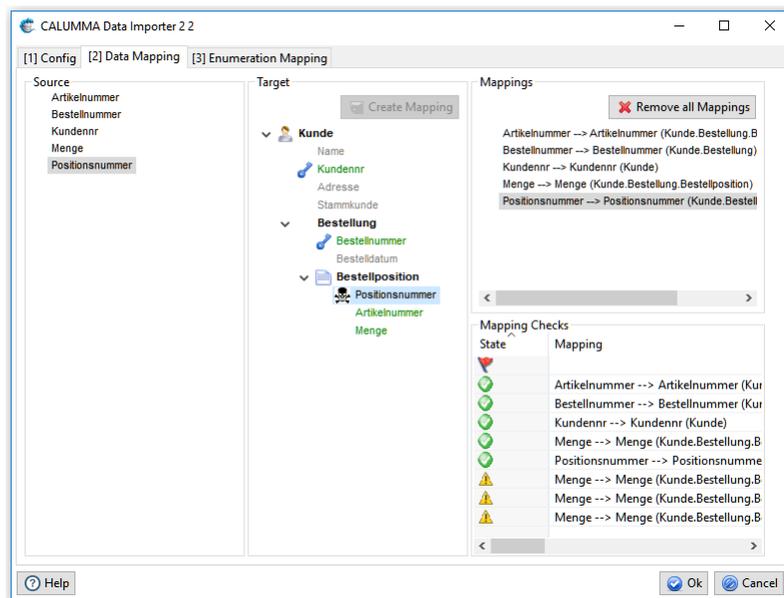


### 3.6 Löschen beim Import

Zuerst muss im ETL-Tool im „Config-Tab“ die Option „Allow Deletes“ aktiviert werden.



Im Data-Mapping-Tab bei der zu löschenden Entity einen Delete-Schlüssel (schwarzer Totenkopf) setzen. Dieser Schlüssel erlaubt auch ein „Create“, also Records werden erstellt wie beim normalen (dunkelblauen) Schlüssel.



Datensätze die sich nicht im Source-File befinden, jedoch aber in der DB, werden gelöscht. Gelöscht wird immer nur innerhalb einer Hierarchie, also es werden in diesem Beispiel nur jene Positionen gelöscht, die zu diesem Kunden und zu dieser Bestellung gehören. Alle anderen Auftragspositionen bleiben erhalten.

Examine preview data

Rows of step: Positionen (22 rows)

#	Bestellnummer	Positionsnummer	Menge	Artikelnummer	Kundenr
1	456567	1	1	234	28345
2	456567	2	2	456	28345
3	4565678	1	1	678	38475
4	4565678	2	3	3454	38475
5	4565678	3	5	34656	38475
6	78789	1	3	23234	348759
7	34576	1	1	56757	349875
8	34576	2	1	345567	349875
9	34576	3	1	12343	349875
10	34577	1	1	4554768	28345
11	34578	1	1	324456	28345
12	34579	1	2	982754	28345
13	34580	1	1	8234795823	28345
14	34581	1	2	943759	28345
15	34582	1	1	9834702	28345
16	34583	1	3	93487523	28345
17	34584	1	2	903845	28345
18	34585	1	1	1234	28345
19	34586	1	1	394875	28345
20	34587	1	1	8723405	28345
21	34588	1	2	2034875	28345
22	34589	1	2	2873450	28345