

tWriteXMLField Example

The tWriteXMLField component is used to load XML into a database column. This post presents a sample job on the topic.

Sometimes, it's helpful to store raw XML in the database. A few scenarios come to mind. One is using a column of raw XML alongside other fields for auditing or system troubleshooting purposes. Another is caching. An XML document may be built up within a database for fast retrieval later.

Job Design

Talend Open Studio's tWriteXMLField works with this structure by mapping the fields of an input source to a field in the output source. Most likely, the output source will be an RDBMS, but could also be a web service.

This example turns a record from a line in a CSV file into an XML document stored in a MySQL column. As the XML is added, an AUTO_INCREMENT id is applied by MySQL.

The source schema is CONTACT_CSV(firstName, city, state). The target schema is CONTACT_XML(contact_xml_id, contact_xml) where contact_xml is a Clob (TEXT) containing an XML document.

The screenshot displays the Talend Open Studio interface. At the top, a job design canvas shows a flow from 'tFileInputDelimited_1' to 'tWriteXMLField_1' and then to a database icon labeled 'contact_xml'. Performance metrics are shown above the components: '2 rows in 0.08s 23.53 rows/s row1 (Main)' for the input and '2 rows in 0.75s 2.66 rows/s row2 (Main)' for the output.

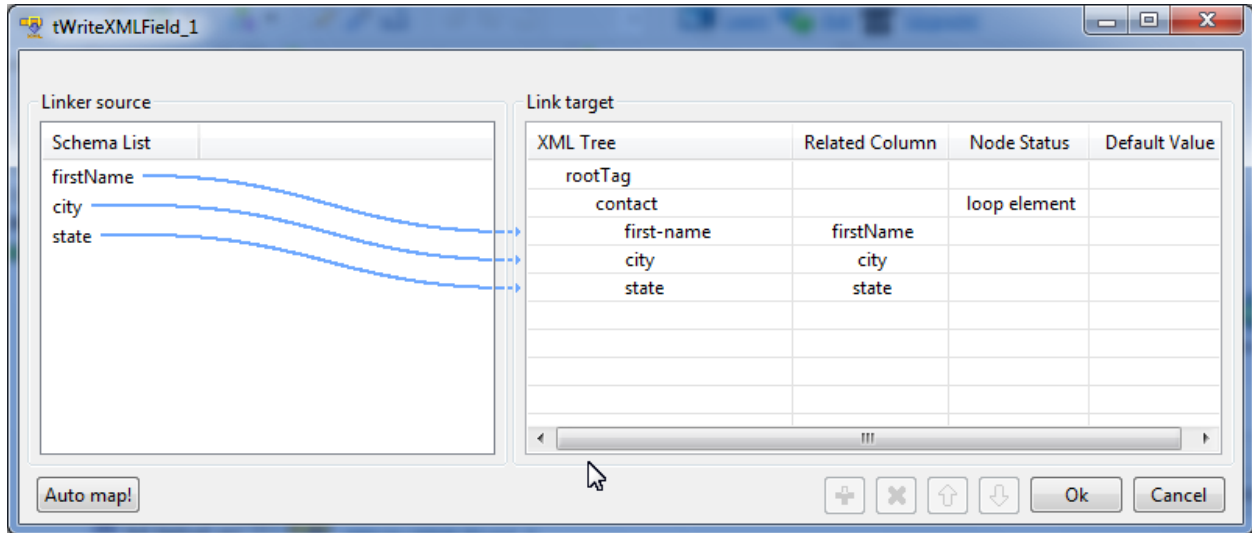
Below the canvas, the 'tWriteXMLField_1' component configuration is visible. The 'Basic settings' tab is active, showing the following configuration:

- Output Column: contact_xml
- Configure Xml Tree: [Icon]
- Schema: Repository | DB (MYSQL):BEKDEK_DEV_Conn - contact_* | Edit schema | Sync columns
- Group by: InputColumn

JOB USING TWRITEXMLFIELD

Configuring XML Component

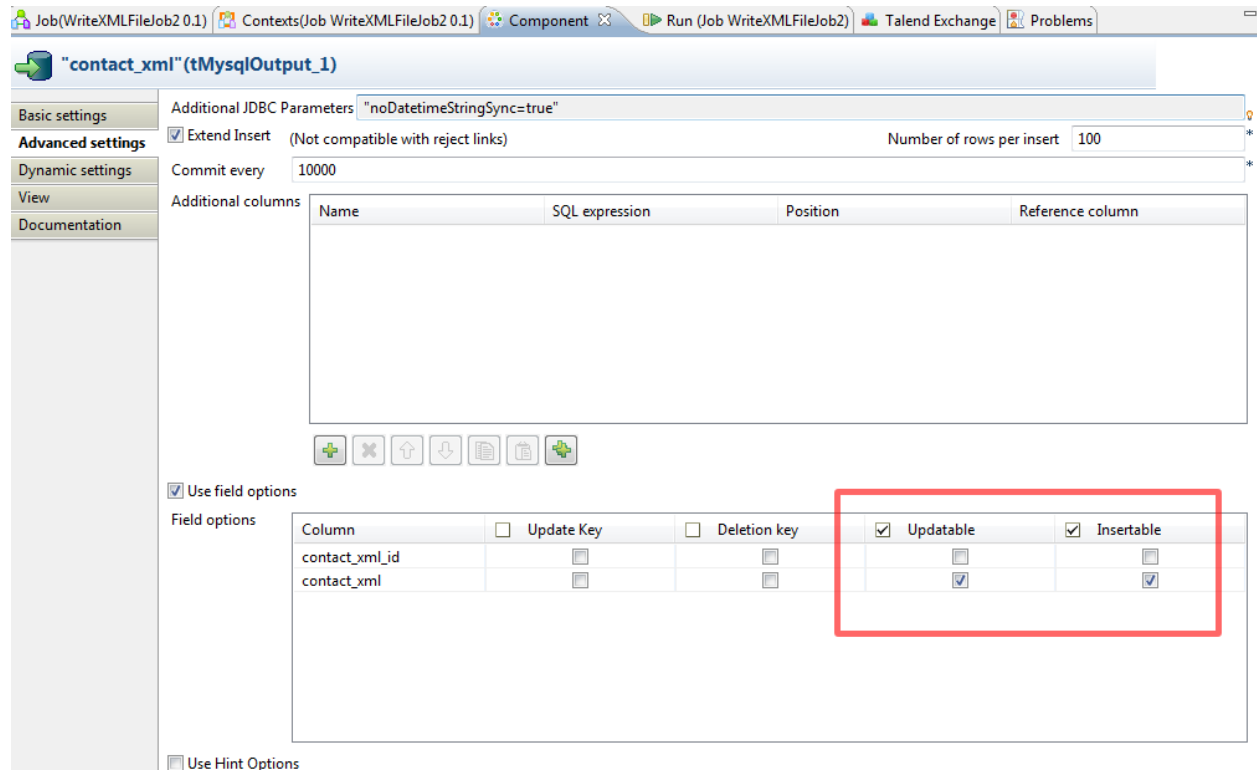
The job in the screenshot is using three components: tFileInputDelimited, tMySQLOutput, and tWriteXMLField. The screenshot shows the configuration of the tWriteField component which directs the resulting XML document to the contact_xml field in MySQL. The following screenshot is the mapping of input to form the XML ("Configure XML Tree" was pressed).



CONFIGURING TWRITEXMLFIELD

MySQL Output

The tMySQLOutput component is a two-field table. Since the id field (contact_xml_id) is handled by MySQL, I use the Advanced settings tab to turn off the insert and update.



Additional JDBC Parameters: "noDatetimeStringSync=true"

Extend Insert (Not compatible with reject links) Number of rows per insert: 100

Commit every: 10000

| Name | SQL expression | Position | Reference column |
|------|----------------|----------|------------------|
|------|----------------|----------|------------------|

Use field options

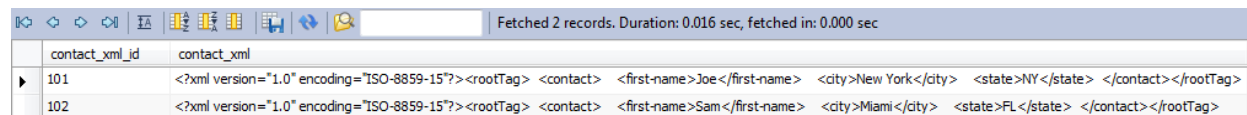
| Column | <input type="checkbox"/> Update Key | <input type="checkbox"/> Deletion key | <input checked="" type="checkbox"/> Updatable | <input checked="" type="checkbox"/> Insertable |
|----------------|-------------------------------------|---------------------------------------|---|--|
| contact_xml_id | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
| contact_xml | <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |

Use Hint Options

TALEND IGNORING ID COLUMN ON INSERT AND UPDATE

Result

The result of running the job on a two-line CSV file is two records in the database, each containing an XML document in the contact_xml field and a unique auto-generated id.



| contact_xml_id | contact_xml |
|----------------|---|
| 101 | <?xml version="1.0" encoding="ISO-8859-15"?><rootTag> <contact> <first-name>Joe</first-name> <city>New York</city> <state>NY</state> </contact></rootTag> |
| 102 | <?xml version="1.0" encoding="ISO-8859-15"?><rootTag> <contact> <first-name>Sam</first-name> <city>Miami</city> <state>FL</state> </contact></rootTag> |

TEXT COLUMN IN MYSQL TABLE

It can be useful to store an XML document in a database for caching or traceability. I've worked with one system that maintains a cached XML document for easy web service retrieval. In the operational environment, this document is updated with Oracle's UPDATEXML function. The column type is Oracle's XMLTYPE. Talend Open Studio can be used to integrate additional data in the XML column or maintain what's there is tWriteXMLField.