

MarkLogic

MarkLogic Kafka Connector

Changes for version 1.2.3

Biju George

7-13-2020

Contents

Summary of this document.....	2
New configurations.....	3
Examples	4
Recipe for URI customization.....	6

Summary of this document

This document describes the changes made to the connector version 1.2.3. Below are the changes / capabilities added

New Capabilities

1. Added two configurable strategies for URI generation apart from the default UUID strategy

New configurations

Below are the new properties for the ID Strategy

Property	Description
<code>ml.id.strategy</code>	<i>JSONPATH</i> : Configure one document path here. The value in this path will become the last part of URI. <i>HASH</i> : Configure one or more comma separated document paths here. The value of last part of the URI would be the MD5 Hashed value of the values concatenated together. Refer the examples for details. This can be usable if you want to use a significant data in the document to form the URI, but do not want to explicitly have the value in the URI.
<code>ml.id.strategy.paths</code>	The comma separated list of document paths for the id strategy. If <code>ml.id.strategy</code> is <i>JSONPATH</i> , only the first path is taken. For <i>HASH</i> , all values are concatenated for generating a HASH value. See the examples below

Examples

Take an example document as below.

```
{
  "Customer": {
    "id": 99102010,
    "FirstName": "Lee",
    "LastName": "Chung",
    "Address": {
      "Line1": "123 Main St.",
      "Line2": "",
      "City": "myCity",
      "State": "CA",
      "Zip": 90002
    },
    "IDs": [
      {
        "type": "SSN",
        "value": "323321111"
      },
      {
        "type": "RESPH",
        "value": "8888887333"
      },
      {
        "type": "MOBILE",
        "value": "8899287333"
      }
    ]
  }
}
```

Scenario 1: Generate URI with the id value.

Configuration:

```
ml.document.uriPrefix=/kafka-data/example/
ml.document.uriSuffix=.json
ml.id.strategy=JSONPATH
ml.id.strategy.paths=/Customer/id
```

URI Generated: /kafka-data/example/99102010.json

Notes:

1. If the path is not valid, the last part of the URI will be blank.
2. The connector will not ensure the uniqueness of the URI generated. So, make sure that the value chosen is unique or it is ok to overwrite.

Scenario2: Generate URI with a hash of id value

Configuration:

```
ml.document.uriPrefix= /kafka-data/example/  
ml.document.uriSuffix=.json  
ml.id.strategy=HASH  
ml.id.strategy.paths=/Customer/id
```

URI Generated: /kafka-data/example/95a033c8257bed4428205f1ee9b487f6.json

Notes:

1. The connector will not ensure the uniqueness of the URI generated. So, make sure that the value chosen is unique or it is ok to overwrite.
2. MD5 Hashing algorithm is used to generate the hash.

Scenario3: Generate URI with a hash of id, AddressZip & SSN

Configuration:

```
ml.document.uriPrefix= /kafka-data/example/  
ml.document.uriSuffix=.json  
ml.id.strategy=HASH  
ml.id.strategy.paths=/Customer/id,/Customer/Address/Zip,/Customer/IDs/0/value
```

URI Generated: /kafka-data/example/d07cb9f262754f56b40f7caecee11ef3.json

Here, d07cb9f262754f56b40f7caecee11ef3 is the HASH of the string 9910201090002323321111

Notes:

1. The connector will not ensure the uniqueness of the URI generated. So, make sure that the value chosen is unique or it is ok to overwrite.
2. See the way of accessing the arrays by index and not by value. So, the order of the array is important if you are using this strategy. In future versions, more flexible options will be provided. For example, suppose non-existing array index is used

```
ml.id.strategy.paths=/Customer/id,/Customer/Address/Zip,/Customer/IDs/5/value
```

The output URI is /kafka-data/example/25955243abb2bb2fd6964aff67fd7c27.json

Here, 25955243abb2bb2fd6964aff67fd7c27 is the HASH of 9910201090002

Recipe for URI customization

If a different strategy which is not available now is required, then `extractId` method of interface

`com.marklogic.client.ext.document.ContentIdExtractor` should be implemented and call the appropriate `extractId` method from the build method in `com.marklogic.client.ext.document.DocumentWriteOperation`. The implementations of current strategies can help in building unique strategies.