Technical Forums (/s/forums)          Knowledge Base (/s/knowledge-base)          Ideas (/s/bridea/acideasULQ__bi          Log in          Q0(

Article

← FME Desktop
(/S/Topic/0TO4Q000000QL9uWAG/Fme-...

## Dynamic Workflows with a Multi-Dataset SchemaScanner

🕐 Aug 4, 2022 • Knowledge

**Product Type**
FME Desktop

**FME Version**
2022.2

**Previous:** Dynamic Workflows with the SchemaScanner (https://community.safe.com/s/article/Dynamic-Workflows-with-the-SchemaScanner) | **Next:** Dynamic
Workflows: Dealing with Schema Drift (https://community.safe.com/s/article/Dynamic-Workflows-Dealing-with-Schema-Drift)

# Introduction

At a basic level, the SchemaScanner is fairly simple to operate. However, there are more complex scenarios that revolve around how the different parameters
are used and the exact values of incoming data.

The Empty attributes parameter controls the schema when incoming data has missing values:
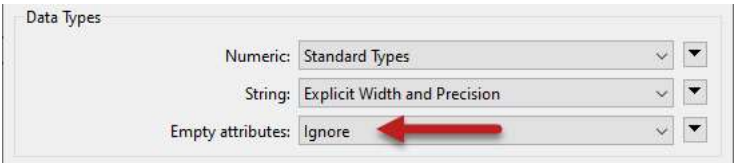


However, this only applies to data where the entire dataset is missing a particular attribute value. In this dataset, for example, every value for MarketType is
null. MergedAddress also has null values, but not for every record:

Given the above parameters, the output schema will not contain the MarketType field, because there are no values at all for it:



However, MergedAddress *is* included, because some records still do have values.

Now, say for example, that we wanted to include MarketType in the output, even though there are no values. There are two alternatives to the Ignore option:



Because there are no data values, the SchemaScanner cannot scan the data to guess the data type. Instead, it can either use a default data type (a varchar of unknown length) or it can trace backwards through the workspace trying to find any clues to the data type.

In this case, the reader itself has that information:



…so the SchemaScanner will use that and create an output schema where MarketType = varchar(15).

# Step-by-step Instructions

The goal here is to add more data to the output from the previous exercise (https://community.safe.com/s/article/Dynamic-Workflows-with-the-SchemaScanner), by merging different sets of source data.
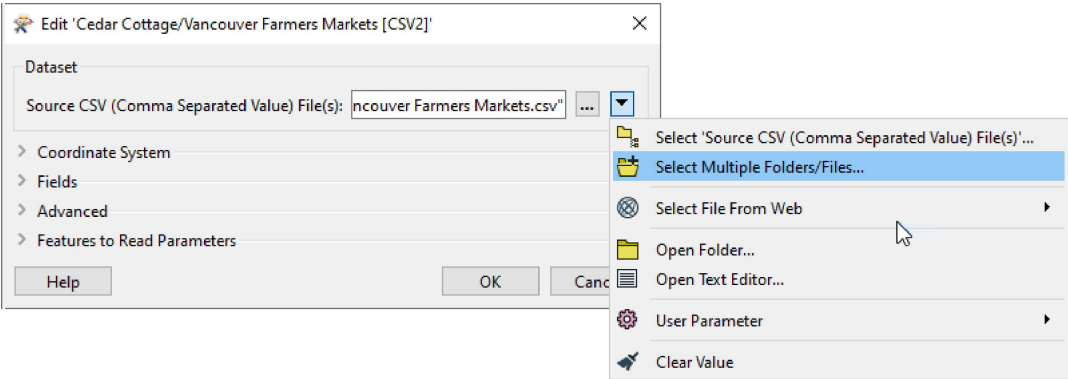
### 1. Open Workspace

Open the workspace from the previous exercise (https://community.safe.com/s/article/Dynamic-Workflows-with-the-SchemaScanner). Now let's add in the Vancouver Farmers Market CSV dataset, keeping the existing Cedar Cottage data. You can do this in one of two ways:

### 1a. Add a New CSV Reader.

Either use Reader > Add Reader from the menubar or drag/drop the CSV file into the workspace canvas. Connect the new feature type to the AttributeManager input port.

### 1b. Edit the Existing CSV Reader

In the Navigator window, double-click the Source CSV parameter for the existing CSV Reader. When it opens, click the drop-down arrow and choose Select Multiple Folders/Files:



In the dialog that opens, click Add Files and choose the file Vancouver Farmers Market CSV. Because the existing reader is dynamic, you won't even have to add a new feature type.

### 2. Run Workspace

Run the workspace and inspect the output. You'll notice that the Vancouver Farmers Market data has no values for MarketType, so their values are Null. Also notice that the data has an extra field called Offerings, which Cedar Cottage is missing:

| | MarketType | MarketName | hs | NumberOfVendors | Offerings |
|---|---|---|---|---|---|
| 1 | Community Food Market | Alternates between 4 locations | | <null> | <null> |
| 2 | Community Food Market | South Vancouver Neighbourhood Ho. | | <null> | <null> |
| 3 | Community Food Market | Brant Villa | ng | 2-3 Tables | <null> |
| 4 | Community Food Market | Cedar Cottage Neighbourhood House | | 2-3 Tables | <null> |
| 5 | Community Food Market | Culloden Court | ng | 2-3 Tables | <null> |
| 6 | Community Food Market | Orchard Park Mobile Market and Tastin | ember | 1 | <null> |
| 7 | Community Food Market | Cedar Cottage Neighbourhood Hous | December | 1 | <null> |
| 8 | Community Food Market | Kensington Community Centre | | 2-3 Tables | <null> |
| 9 | Community Food Market | Orchard Park | y | <null> | <null> |
| 10 | Community Food Market | Culloden Court Mobile Market and Ta | -December | 1 | <null> |
| 11 | Community Food Market | Brant Villa Mobile Market | December | 1 | <null> |
| 12 | <null> | Hastings Park | pr | 33 | produce,meat,seafood,cheese,organics,prepared foods |
| 13 | <null> | Mount Pleasant Market | t | 25 | produce,meat,seafood,cheese,organics,prepared foods |
| 14 | <null> | Kitsilano Farmers Market | Oct | 51 | produce,meat,seafood,cheese,organics,prepared foods |
| 15 | <null> | Main St Station Farmers Market | ct | 28 | produce,meat,seafood,cheese,organics,prepared foods |

These fields were included in the output CSV, even though not all data possessed them. That's because the SchemaScanner creates an inclusive schema. i.e. if

only one feature possesses an attribute, then it will be included in the output.

### 3. Create Different Outputs for Each Day

Now let's create a different file for each day of the week; i.e. we wish to divide food markets into different outputs depending on which day they take place.
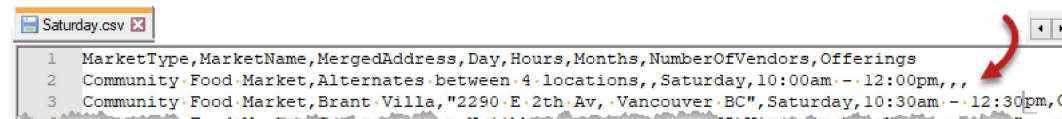
Open the parameters dialog for the writer feature type.

Change the CSV File Name parameter from fme_feature_type to the Day attribute by clicking on the drop-down arrow.
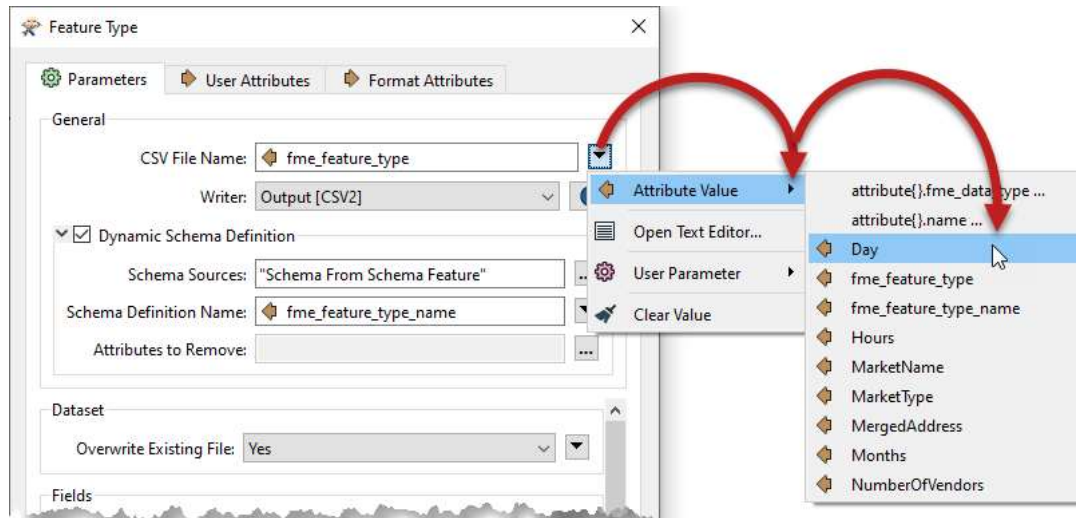


Re-run the workspace and inspect the outputs in a Text Editor.

In the Saturday and Sunday files, some fields have null values, but the fields still exist because not every record is null:



In the Wednesday and Thursday files, some fields are entirely null. However, the fields still exist because there is only one output schema and every output dataset is using it.
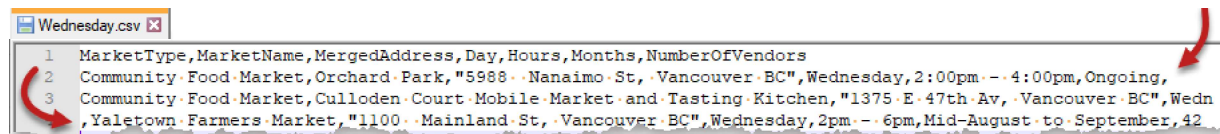


Let's see how we could give each output a different schema.

### 4. Update SchemaScanner to Group Features

View the SchemaScanner parameters. Enable Group Processing and under Group By, choose the Day attribute.

Re-run the workspace. Notice that there are now four schema features, one per day of the week:



Notice that fme_feature_type_name - the attribute that sets the name of the schema - is what defines the difference between each feature.
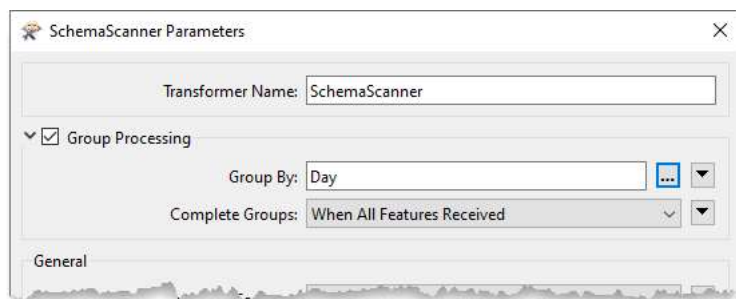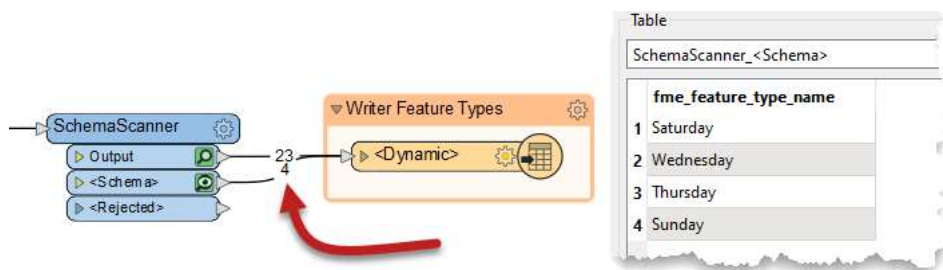
Inspect the output. If a dataset was missing an entire attribute, then that attribute is now absent in the output. That's because each dataset has its own schema, and the SchemaScanner is set to Ignore Empty Attributes.

## 5. Output Individual Schemas

Finally, let's say we wanted to give each output its own schema, but we didn't want empty attributes to be ignored. How would we do that?

View the SchemaScanner parameters again. Change the Empty Attributes parameter from "Ignore" to "Interpret Upstream Schema (Advanced)". Now FME will include those attributes, obtaining the information to create the schema from wherever it can find that "upstream" in the workspace.

Once again, run the workspace and inspect the Schema features, in particular the one for Thursday markets.



Previously the Thursday market did not have a MarketType attribute in its schema. Now it does, even though all of the values are empty. The type is fme_varchar(22) because FME could determine that from the reader schema.

## Continue to the next article: Dynamic Workflows: Dealing with Schema Drift (https://community.safe.com/s/article/Dynamic-Workflows-Dealing-with-Schema-Drift)

## Data Attribution

The data used here originates from open data made available by the City of Vancouver (https://opendata.vancouver.ca/pages/home/), British Columbia. It contains information licensed under the Open Government License - Vancouver.

---

**First Published Date**
7/29/2022, 5:14 PM

---

**Last Published Date**
8/4/2022, 3:10 PM

---

Authoring
(/s/topic/0TO4Q000000...

FME Desktop
(/s/topic/0TO4Q000000...

Sort by:

Latest Posts ▾

LizAtSafe (/s/profile/0050c00000CeGV0AAN) (Employee) published a new version of this Knowledge.
August 4, 2022 at 3:10 PM (/s/feed/0D54Q00009hQzPpSAK)

👍 Like                                                    💬 Comment

Log In to Comment

mark2atsafe (/s/profile/00530000003Tn36AAC) (Employee) published this new Knowledge.
July 29, 2022 at 5:14 PM (/s/feed/0D54Q00009gck1PSAQ)

👍 Like                                                    💬 Comment

Log In to Comment

Follow

Files (1) (/s/relatedlist/ka14Q000001DWywQAG/AttachedContentDocuments)

ZIP    4-WorkflowWithAMultiDatasetSchemaScanner
       Aug 4, 2022  •  64KB  •  zip

View All
(/s/relatedlist/ka14Q000001DWywQAG/AttachedContentDocuments)

## Related Articles

Tutorial: Dynamic Workflows (/s/article/dynamic-workflow-tutorial-introduction)

Dynamic Workflows: Advanced Example - Modifying the Schema Feature (/s/article/dynamic-workflows-advanced-example-modifying-the-s)

Dynamic Workflows: Destination Schema Derived from Multiple Sources (/s/article/dynamic-workflows-destination-schema-derived-from)

Dynamic Workflows: Destination Schema as a Mirror Image of the Source Dataset (/s/article/dynamic-workflow-tutorial-destination-schema-as-a)

Dynamic Workflows with the SchemaScanner (/s/article/Dynamic-Workflows-with-the-SchemaScanner)

**SAFE SOFTWARE**®
(https://safe.com)

Getting Started
(../s/topic/0TO4Q000000QKioWAG/welcome)

Forums (../s/forums/)

Knowledge Base (../s/knowledge-base/)

Ideas
(../s/bridea/acideasULT__brIdea__c/00B4Q000000AwHEYUAA)

Groups
(../s/group/CollaborationGroup/00Ba000000A0BxJEAV)

Support (../s/support/)

Feedback
(../s/idea/edit?retURL=LkoFWpziDYaWQkL78)

**Register / Log In (/s/login/)**

© Safe Software Inc | Legal (https://www.safe.com/legal/)

(http://twitter.com/SafeSoftware/))

**Land Acknowledgement —**

Safe Software respectfully acknowledges that we live, learn and work on the traditional and unceded territories of the Kwantlen, Katzie, and Semiahmoo First Nations.