

Schema compatibility & evolution

1. Schema registry (default compatibility (backward))

Confluent Avro schema registry is being used by producer & consumer of the messages. A valid compatibility setting has been identified & used. Various possible use cases are considered, experimented & documented here. All schemas used are default compatibility (backward)

2. Input and Job Used

Used input table rate_card and unit_length

Flink Job: ProcessRatecardConformance(Indcdcadsrcrd_ratecard, Indcdcadsvntry_unitlength)

3. Use Cases

3.1. Use case -1 -?A new field/column is added at the source advisor system; but that column is not used by the consumer.

3.1.1. Scenario-1: the consumer process the new data??

Ran producer to produce the data

Ran ProcessRatecardConformance and consumed data produced and process worked as before with no new column in final result of conformance_target_consr_rtcd but new column was there in input data for topic Indcdcadsrcrd_ratecard)

3.1.2. Scenario-2: the consumer re-process the entire data?

Ran ProcessRatecardConformance and consumed data produced from couple of days ago till date and process worked as before with no new column in final result of conformance_target_consr_rtcd

3.2. Use case -2 -?A new field/column is added at the source advisor system; that column is being used by the consumer

3.2.1. Scenario-1: The consumer process that data that is produced after the change

Added new column

```
{
  "name" : "RATE_CARD_LONG_NM",
  "type" : [ "null", "String" ],
  "default": null
}
```

Updated CREATE TABLE Indcdcadsrcrd_ratecard to include new column

Updated CREATE TABLE conformance_target_consr_rtcd to include new column

Ran producer to produce the data

Ran ProcessRatecardConformance and consumed data produced and process worked with new column in final result of conformance_target_consr_rtcd and new column was there in data for topic Indcdcadsrcrd_ratecard).

New column showed expected data

3.2.2. Scenario-2: The consumer re-process the entire data

Ran ProcessRatecardConformance with option to start from beginning and consumed data produced and process worked new column in final result of conformance_target_consr_rtcd and new column was there in data for topic Indcdcadsrcrd_ratecard).

New column showed expected data for data produced after adding new column in the schema with new data in new column

New column showed null value for data produced after adding new column in the schema with no data in new column and default of null

3.3. Use case-3 - A field at the source advisor system is deleted; but that field is not used by the consumer.

Only works??with Optional fields with default value

3.3.1. Scenario-1: The consumer process that data that is produced after the change

Changes rate_card topic DIVISION_ID column schema to make it optional with default value. Produced some records with no value for DIVISION_ID

```
{
  "name": "DIVISION_ID",
  "type": [ "null", "long" ],
  "default": null
}
```

Updated?ProcessRatecardConformance to not use?DIVISION_ID

Ran ProcessRatecardConformance and consumed data produced and process worked as before with expected data in conformance_target_consr_rtcd

3.3.2. Scenario-2: The consumer re-process the entire data

Changes rate_card topic DIVISION_ID column schema to make it optional with default value. Produced some records with no value for DIVISION_ID

```
{
  "name": "DIVISION_ID",
  "type": [ "null", "long" ],
  "default": null
}
```

Updated?ProcessRatecardConformance to not use?DIVISION_ID

Ran ProcessRatecardConformance and consumed data produced from few days ago? and process worked as before with expected data in conformance_target_consr_rtcd

3.4. Use case-4 - A field at the source advisor system is deleted; that field is used by the consumer

Only works??with Optional fields with default value

3.4.1. Scenario-1: The consumer process that data that is produced after the change

Changes rate_card topic DIVISION_ID column schema to make it optional with default value. Produced some records with no value for DIVISION_ID.

Updated?ProcessRatecardConformance to use COALESCE

Ran ProcessRatecardConformance and consumed data produced and process worked? as expected data using the default value of DIVISION_ID in conformance_target_consr_rtcd

3.4.2. Scenario-2: The consumer re-process the entire data

Changes rate_card topic DIVISION_ID column schema to make it optional with default value. Produced some records with no value for DIVISION_ID.

Updated?ProcessRatecardConformance to use COALESCE

Ran ProcessRatecardConformance and consumed data produced starting from few days ago? and process worked? as expected data using either the default value or the actual value iDIVISION_ID? n the message? in conformance_target_consr_rtcd

3.5. Use case-5 - A field at the source advisor system is deleted; but that field is not used by the consumer (Non optional fields)

From backward compatibility point of view its not supported

Other option :? 1. Consume all the existing data.? 2. Update consumer schema. 3.Update producer schema and then produce the data? 4.?Consume the data using Job.? Re-processing will not be supported

3.6. Use case-6 - A field at the source advisor system is deleted; but that field is used by the consumer (Non optional fields)

From backward compatibility point of view its not supported

Other option :? 1. Consume all the existing data.? 2. Update consumer schema. 3.Update producer schema and then produce the data? 4. Update Flink Job code to remove the deleted fields. 5.?Consume the data using Job.? Re-processing will not be supported

3.7. Use case-7 - A field at source advisor system is updated; but the column is not used by the consumer.

From backward compatibility point of view its not supported

Other option: Similar to delete non optional fields



Error updating the subject with this schema: Schema being registered is incompatible with an earlier schema for subject "Indcdcadstcrd_ratecard-value"; error code: 409

3.8. Use case-8 - A field at source advisor system is updated; that column is used by the consumer.

From backward compatibility point of view its not supported

Other option: Similar to delete non optional fields



Error updating the subject with this schema: Schema being registered is incompatible with an earlier schema for subject "Indcdcadstcrd_ratecard-value"; error code: 409

4. Schema registry (Full transitive) WIP

Confluent Avro schema registry is being used by producer & consumer of the messages. A valid compatibility setting has been identified & used. Various possible use cases are considered, experimented & documented here. All schema used are FULL_TRANSITIVE

The scope of this test is to create multiple versions of schemas by adding a new column (optional) in each schema

```
# Update compatibility requirements globally
$ curl -X PUT -H "Content-Type: application/vnd.schemaregistry.v1+json" \
  --data '{"compatibility": "FULL_TRANSITIVE"}' \
  http://localhost:8081/config
```

5. Input and Job Used

Used input table rate_card and unit_length

Flink Job: ProcessRatecardConformance(Indcdcadstcrd_ratecard, Indcdcadstcrd_ratecard)

5.1. Use case -1 -?A new field/column is added at the source advisor system; but that column is not used by the consumer. Using V1, V2, V3 schemas

V2?? Add a new optional column?RATE_CARD_NM_DESC in?Indcdcadstcrd_ratecard

V3 ? Add a new optional column?RATE_CARD_NOTES? ? in?Indcdcadstcrd_ratecard

5.1.1. Scenario-1: the consumer process the new data??

Ran ProcessRatecardConformance and then consumed V1,? V2, V3 and when version were updated.? Job run as before with no new columns

5.1.2. Scenario-2: the consumer re-process the entire data?

Ran ProcessRatecardConformance and then consumed V1,? V2, V3 data? ?Job run as before with no new columns

5.2. Use case -2 -?A new field/column is added at the source advisor system; that column is being used by the consumer.?

5.3. Using V1, V2, V3 schemas

V2?? Add a new optional column?RATE_CARD_NM_DESC in?Indcdcdsrtcrd_ratecard

V3 ? Add a new optional column?RATE_CARD_NOTES? ? in?Indcdcdsrtcrd_ratecard

5.3.1. Scenario-1:? The consumer process that data that is produced after the change

Ran ProcessRatecardConformance and then consumed V1,? V2, V3 and then updated the job to include new fields as and when versions were updated.? Job run as before with new columns in the outputs as per the versions added

5.3.2. Scenario-2: The consumer re-process the entire data

Ran ProcessRatecardConformance and then consumed V1,? V2, V3. Job run as before with new columns in the outputs with values? as published in the messageor default values as per the versions added