

Talend For DI & BD Certification Training Certification Project

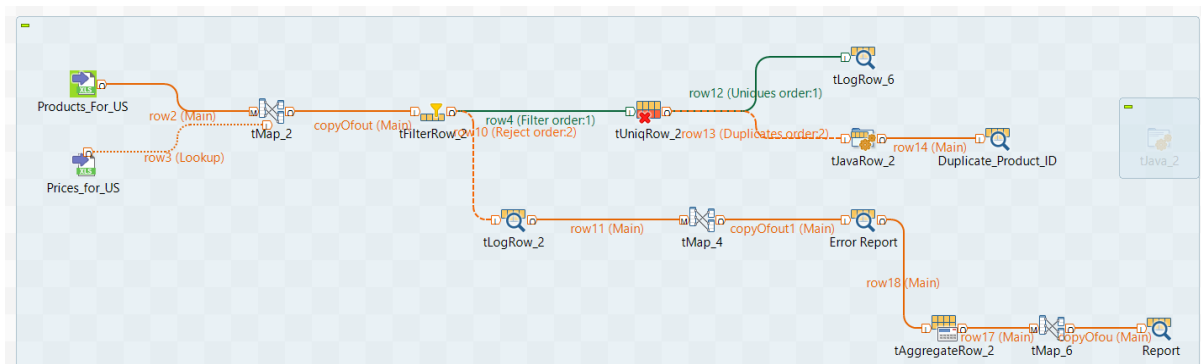
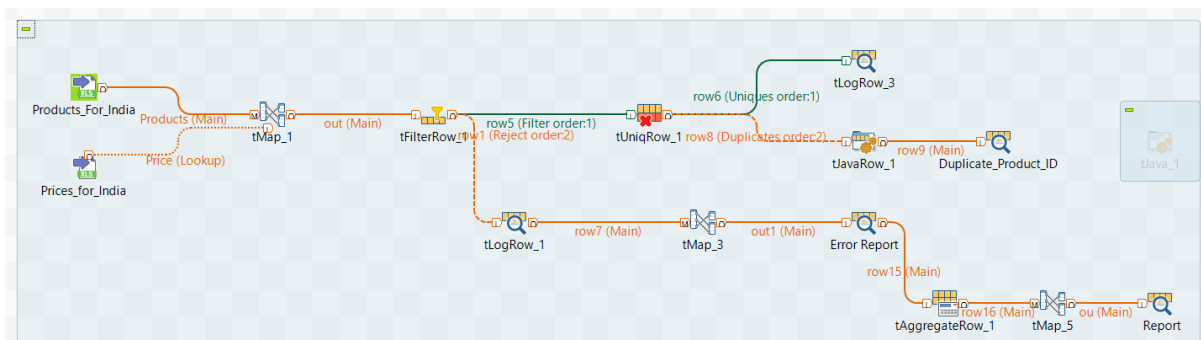
Problem Statement:

In any E-commerce website when a new product has to be added for sale, it has to adhere to many business Rules. (For multiple countries i.e., product attributes change w.r.t country Ex: Language changes as per country)

Consider a case where we have to upload a large number of new products for each country to E-commerce website. Below are the problems:

- i. If a product did not adhere to any one business rule it is rejected, but product owner is not able to figure out for which business rule it is failing to correct the data
- ii. If product owner is correcting one error by looking at the error report and again uploads data, it may fail for next business rule i.e. all possible errors are not captured in one go.
- iii. Management is not able to have a big picture of statistics i.e. for each country how many products passed and how many failed. So that they can concentrate on the country which has less data quality for product data.

Problem Solution:



1. Input Files will be provided country wise i.e. each country will have separate input file.

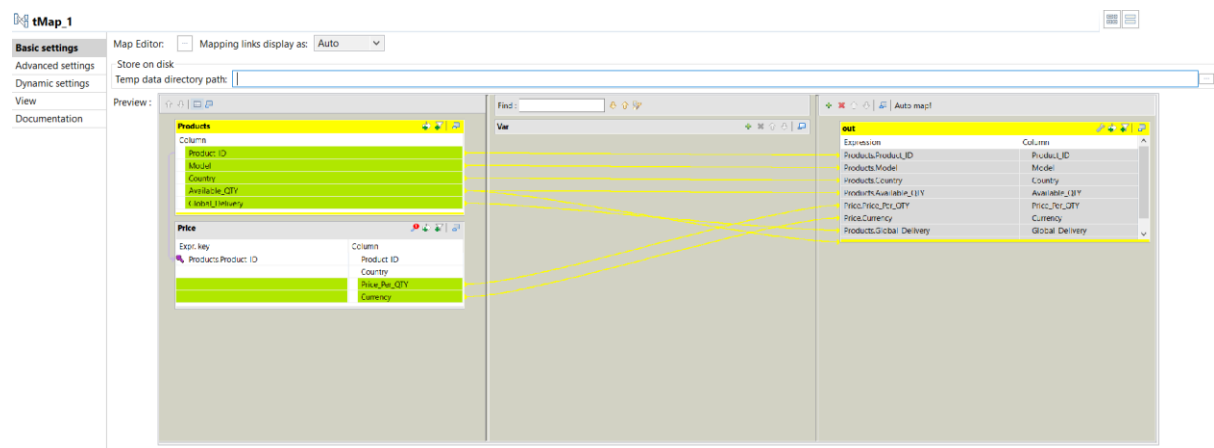
Products_For_India.csv

Product ID	Model	Country	Available QTY	Global Delivery
1	122345	IN	34	NO
2	545484	IN	56	YES
3	765434	IN	12	YES
	123443	IN	56	NO
5	674323	IN	19	YES
6	323245	IN	20	YES
7	323245	IN	0	NO
8		IN	56	YES
9	674323	IN	222	YES
10	323245	IN	0	NO
11	545484	IN	554	YES
12	765434	IN	118	YES
13	765434	IN	582	NO
2	122345	IN	137	YES
15	765434		472	YES
16	765434	IN	106	NO
17	765434	IN	235	YES
18	765434		605	YES
19	765434	IN	813	NO
20	123443	IN	389	YES
21	674323	IN	435	YES
22	323245	IN	546	NO
23	765434	IN	406	YES
24	674323	IN	579	YES
25	674323	IN	521	NO
26	674323	IN	898	YES
27	123443	IN	307	YES
28	674323	IN	63	NO
29	323245	IN	797	YES

Prices_For_India.csv

Product ID	Country	Price Per C	Currency
1	IN	803	INR
2	IN	619	INR
3	IN	524	INR
4	IN	600	INR
5	IN		INR
6	IN	697	INR
7	IN	731	INR
8	IN	477	INR
9	IN	246	INR
10	IN	710	INR
11	IN	824	INR
12	IN	462	INR
13	IN	647	
14	IN	518	INR
15	IN	473	INR
16	IN	577	INR
17	IN	335	INR
18	IN	425	INR
19	IN	410	INR
20	IN	985	INR
21	IN		INR
22	IN	401	INR
23	IN	511	INR
24	IN	909	INR
25	IN	346	INR
26	IN	718	INR
27	IN	424	INR
28	IN	788	INR
29	IN	645	INR

Staging data Process using tMap Component



India_Products_Staging

```
[statistics] connecting to socket on port 3718
[statistics] connected
```

Products_Staging						
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery
1	122345	IN	34	803	INR	NO
2	545484	IN	56	619	INR	YES
3	765434	IN	12	524	INR	YES
null	123443	IN	56	null	null	NO
5	674323	IN	19	null	INR	YES
6	323245	IN	20	697	INR	YES
7	323245	IN	0	731	INR	NO
8	null	IN	56	477	INR	YES
9	674323	IN	222	246	INR	YES
10	323245	IN	0	710	INR	NO
11	545484	IN	616	824	INR	YES
12	765434	IN	166	462	INR	YES
13	765434	IN	707	647	null	NO
2	122345	IN	314	619	INR	YES
15	765434	null	990	473	INR	YES
16	765434	IN	967	577	INR	NO
17	765434	IN	823	335	INR	YES
18	765434	null	575	425	INR	YES
19	765434	IN	167	410	INR	NO
20	123443	IN	195	985	INR	YES
21	674323	IN	76	null	INR	YES
22	323245	IN	680	401	INR	NO
23	765434	IN	798	511	INR	YES
24	674323	IN	176	909	INR	YES
25	674323	IN	929	346	INR	NO
26	674323	IN	790	718	INR	YES
27	123443	IN	981	424	INR	YES
28	674323	IN	144	788	INR	NO
29	323245	IN	950	645	INR	YES
30	323245	IN	406	370	INR	YES
31	null	IN	167	438	INR	NO
32	123443	IN	848	201	INR	YES
33	674323	IN	67	479	INR	YES
34	323245	IN	130	497	INR	NO
35	765434	IN	686	254	INR	YES

Products_For_US.csv

Product ID	Model	Country	Available QTY	Global Delivery
1	122345	US	34	NO
2	545484	US	56	YES
3		US	12	YES
4	123443	US	56	NO
5	674323	US	19	YES
6	323245	US	0	YES
7	323245	US	123	NO
8	123567	US	56	YES
9	674323	US	222	YES
10	323245	US	567	NO
11	545484	US	83	YES
12	765434	US	863	YES
13	765434	US	520	NO
14	122345	US	160	YES
15	765434		706	YES
16	765434	US	925	NO
17	765434	US	0	YES
18	765434	US	531	YES
19	765434	US	995	NO
20	123443	US	959	YES
21	674323	US	889	YES
22	323245	US	927	NO
23	765434	US	988	YES
24	674323	US	755	YES
25	674323	US	483	NO
26	674323	US	899	YES
27	123443	US	637	YES
28	674323	US	438	NO
29	323245	US	262	YES

Prices_For_US.csv

Product ID	Country	Price Per QTY	Currency
1	US	219	USD
2	US	756	USD
3	US	494	USD
4	US	356	USD
5		708	USD
6	US	729	USD
7	US	677	USD
8	US	439	USD
9	US	544	USD
10	US	402	USD
11	US	315	USD
12	US	922	USD
13	US	452	USD
14	US		USD
	US	931	USD
16	US	311	USD
17	US	578	USD
18	US	957	USD
19	US	270	USD
20	US	858	USD
21	US	474	USD
22	US	860	USD
23	US	632	USD
24	US	657	USD
25	US	699	USD
26	US	451	USD
27	US	565	USD
28	US	621	USD
29	US	849	USD

US_Products_Staging

Products_Staging						
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery
1	122345	US	34	219	USD	NO
2	545484	US	56	756	USD	YES
3	null	US	12	494	USD	YES
4	123443	US	56	356	USD	NO
5	674323	US	19	708	USD	YES
6	323245	US	0	729	USD	YES
7	323245	US	123	677	USD	NO
8	123567	US	56	439	USD	YES
9	674323	US	222	544	USD	YES
10	323245	US	567	402	USD	NO
11	545484	US	81	315	USD	YES
12	765434	US	381	922	USD	YES
13	765434	US	232	452	USD	NO
14	122345	US	134	null	USD	YES
15	765434	null	300	null	null	YES
16	765434	US	926	311	USD	NO
17	765434	US	0	578	USD	YES
18	765434	US	955	957	USD	YES
19	765434	US	389	270	USD	NO
20	123443	US	430	858	USD	YES
21	674323	US	337	474	USD	YES
22	323245	US	379	860	USD	NO
23	765434	US	776	632	USD	YES
24	674323	US	558	657	USD	YES
25	674323	US	223	699	USD	NO
26	674323	US	638	451	USD	YES
27	123443	US	587	565	USD	YES
28	674323	US	852	621	USD	NO
29	323245	US	946	849	USD	YES
30	323245	US	403	273	USD	YES
31	123456	US	899	544	USD	NO
32	123443	US	299	488	USD	YES
33	674323	US	693	216	USD	YES
34	123456	US	730	656	USD	NO
35	765434	US	252	534	USD	YES
36	674323	US	237	735	USD	YES
36	323245	US	730	735	USD	NO

2. Once staging the data is ready, apply the below business Rules.

A. Product ID should be unique for each country. (i.e. same country cannot have two products with same product id, however same product can exist for two Countries). Reject the product if it is duplicate for same country and log error Message. (Implement knowledge of unique/ filter components)

Basic settings
Schema: Built-In Edit schema Sync columns

Advanced settings
Dynamic settings
View
Documentation

Unique key

Column	<input type="checkbox"/> Key attribute	<input type="checkbox"/> Case Sensitive
Product_ID	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Model	<input type="checkbox"/>	<input type="checkbox"/>
Country	<input type="checkbox"/>	<input type="checkbox"/>
Available_QTY	<input type="checkbox"/>	<input type="checkbox"/>
Price_Per_QTY	<input type="checkbox"/>	<input type="checkbox"/>
Currency	<input type="checkbox"/>	<input type="checkbox"/>
Global_Delivery	<input type="checkbox"/>	<input type="checkbox"/>
OUT_OF_STOCK	<input type="checkbox"/>	<input type="checkbox"/>

Basic settings

Advanced settings

Dynamic settings

View

Documentation

Schema

Built-In

Edit schema

Sync columns

Code

Generate code

```
//Code generated according to input schema and output schema
output_row.Product_ID = input_row.Product_ID;
output_row.Country = input_row.Country;
output_row.Business_Rule = "A";
output_row.Attribute = "Product ID";
output_row.Reason = "Duplicate Product ID";
```

Duplicate_Product_ID				
Product_ID	Country	Business_Rule	Attribute	Reason
2	IN	A	Product ID	Duplicate Product ID

Duplicate_Product_ID				
Product_ID	Country	Business_Rule	Attribute	Reason
36	US	A	Product ID	Duplicate Product ID

B. Reject the product if either of the below fields are NULL or empty on invalid by logging separate error message for each field. i. Product ID ii. Model. iii. Country iv. Price v. Currency.

Basic settings

Advanced settings

Dynamic settings

View

Documentation

Schema

Built-In

Edit schema

Sync columns

Logical operator used to combine conditions

And

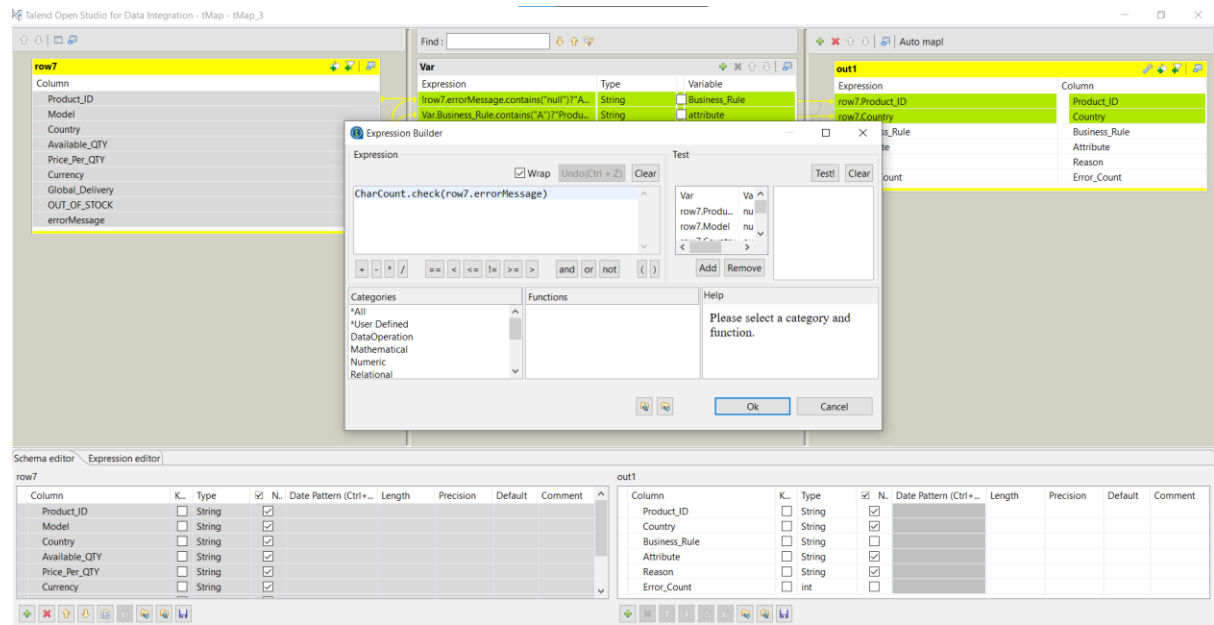
Conditions	InputColumn	Function	Operator	Value
	Product_ID	Empty	Not equal to	null
	Model	Empty	Not equal to	null
	Country	Empty	Not equal to	null
	Price_Per_QTY	Empty	Not equal to	null
	Currency	Empty	Not equal to	null

Use advanced mode

tLogRow_1								
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery	OUT_OF_STOCK	errorMessage
null	123443	IN	56	null	null	NO	FALSE	Product_ID =null failed Price_Per_QTY =null failed Currency =null failed
5	674323	IN	19	null	INR	YES	FALSE	Price_Per_QTY =null failed
8	null	IN	56	477	INR	YES	FALSE	Model =null failed
13	765434	IN	707	null	NO	YES	FALSE	Currency =null failed
15	765434	null	990	473	INR	YES	FALSE	Country =null failed
18	765434	null	575	425	INR	YES	FALSE	Country =null failed
21	674323	IN	76	null	INR	YES	FALSE	Price_Per_QTY =null failed
31	null	IN	167	438	INR	NO	FALSE	Model =null failed
43	323245	IN	558	544	null	NO	FALSE	Currency =null failed

tLogRow_2								
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery	OUT_OF_STOCK	errorMessage
3	null	US	12	494	USD	YES	FALSE	Model =null failed
14	122345	US	134	null	USD	YES	FALSE	Price_Per_QTY =null failed
15	765434	null	300	null	YES	FALSE	FALSE	Country =null failed Price_Per_QTY =null failed Currency =null failed
54	765434	US	794	null	USD	YES	FALSE	Price_Per_QTY =null failed
117	765434	US	485	942	USD	YES	FALSE	Currency =null failed
141	674323	US	577	null	USD	YES	FALSE	Price_Per_QTY =null failed
147	323245	US	743	547	null	YES	FALSE	Currency =null failed

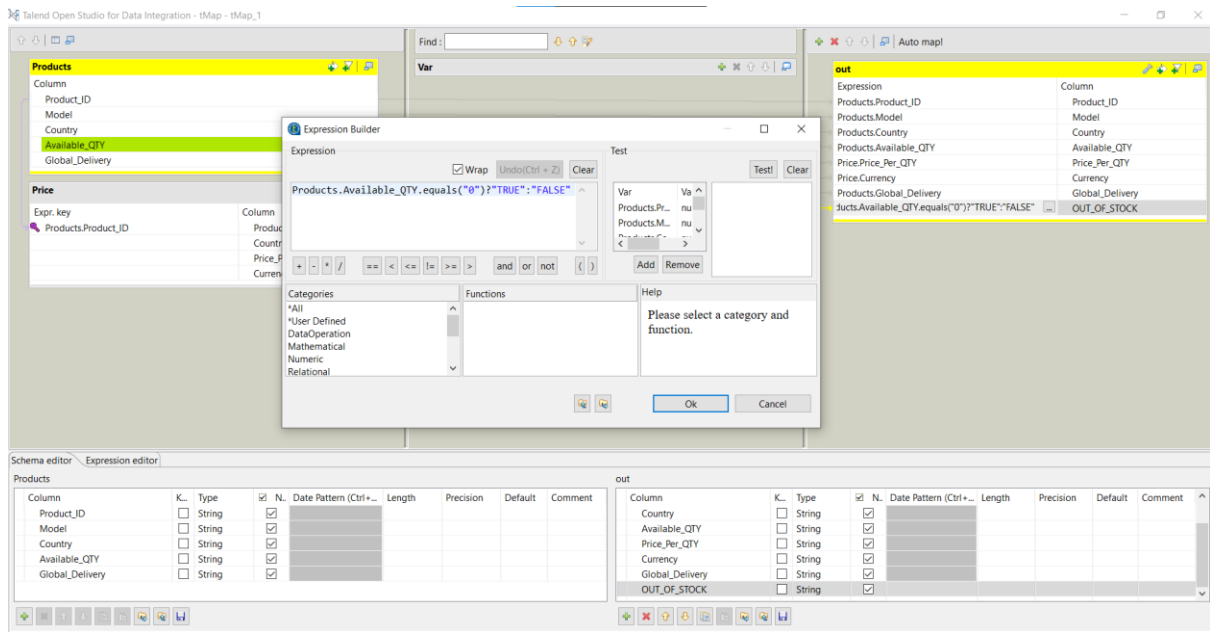
3. Error Report Format should be as below and should be loaded.



Error Report						
Product_ID	Country	Business_Rule	Attribute	Reason		Error_Count
null	IN	B	ID Price Currency	ID is null Price is null Currency is null		3
5	IN	B	Price	Price is null		1
8	IN	B	Model	Model is null		1
13	IN	B	Currency	Currency is null		1
15	null	B	Country	Country is null		1
18	null	B	Country	Country is null		1
21	IN	B	Price	Price is null		1
31	IN	B	Model	Model is null		1
43	IN	B	Currency	Currency is null		1

Error Report						
Product_ID	Country	Business_Rule	Attribute	Reason		Error_Count
3	US	B	Model	Model is null		1
14	US	B	Price	Price is null		1
15	null	B	Country Price Currency	Country is null Price is null Currency is null		3
54	US	B	Price	Price is null		1
117	US	B	Currency	Currency is null		1
141	US	B	Price	Price is null		1
147	US	B	Currency	Currency is null		1

**4. On the data which has successfully passed the business rules, perform the following actions:
Introduce a new column OUT OF STOCK and populate the value for that field as TRUE if 'Available QTY' from input file is 0 else populate the field as FALSE.**



INDIA_PRODUCTS

tLogRow_3							
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery	OUT_OF_STOCK
1	122345	IN	34	803	INR	NO	FALSE
2	545484	IN	56	619	INR	YES	FALSE
3	765434	IN	12	524	INR	YES	FALSE
6	323245	IN	20	697	INR	YES	FALSE
7	323245	IN	0	731	INR	NO	TRUE
9	674323	IN	222	246	INR	YES	FALSE
10	323245	IN	0	710	INR	NO	TRUE
11	545484	IN	616	824	INR	YES	FALSE
12	765434	IN	166	462	INR	YES	FALSE
16	765434	IN	967	577	INR	NO	FALSE
17	765434	IN	823	335	INR	YES	FALSE
19	765434	IN	167	410	INR	NO	FALSE
20	123443	IN	195	985	INR	YES	FALSE
22	323245	IN	680	401	INR	NO	FALSE
23	765434	IN	798	511	INR	YES	FALSE
24	674323	IN	176	909	INR	YES	FALSE
25	674323	IN	929	346	INR	NO	FALSE
26	674323	IN	790	718	INR	YES	FALSE
27	123443	IN	981	424	INR	YES	FALSE
28	674323	IN	144	788	INR	NO	FALSE
29	323245	IN	950	645	INR	YES	FALSE
30	323245	IN	406	370	INR	YES	FALSE
32	123443	IN	848	201	INR	YES	FALSE
33	674323	IN	67	479	INR	YES	FALSE
34	323245	IN	130	497	INR	NO	FALSE
35	765434	IN	686	254	INR	YES	FALSE
36	674323	IN	172	342	INR	YES	FALSE
37	323245	IN	436	431	INR	NO	FALSE
38	674323	IN	581	349	INR	YES	FALSE
39	323245	IN	794	617	INR	YES	FALSE
40	122345	IN	708	912	INR	NO	FALSE
41	545484	IN	285	469	INR	YES	FALSE
42	765434	IN	641	525	INR	YES	FALSE
44	1232345	IN	746	506	INR	YES	FALSE
45	123443	IN	433	430	INR	YES	FALSE
46	674323	IN	694	631	INR	NO	FALSE
47	323245	IN	899	255	INR	YES	FALSE
48	545484	IN	552	602	INR	YES	FALSE
49	765434	IN	222	222	INR	YES	FALSE

US_PRODUCTS

tLogRow_6							
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery	OUT_OF_STOCK
1	122345	US	34	219	USD	NO	FALSE
2	545484	US	56	756	USD	YES	FALSE
4	123443	US	56	356	USD	NO	FALSE
5	674323	US	19	708	USD	YES	FALSE
6	323245	US	0	729	USD	YES	TRUE
7	323245	US	123	677	USD	NO	FALSE
8	123567	US	56	439	USD	YES	FALSE
9	674323	US	222	544	USD	YES	FALSE
10	323245	US	567	402	USD	NO	FALSE
11	545484	US	81	315	USD	YES	FALSE
12	765434	US	381	922	USD	YES	FALSE
13	765434	US	232	452	USD	NO	FALSE
16	765434	US	926	311	USD	NO	FALSE
17	765434	US	0	578	USD	YES	TRUE
18	765434	US	955	957	USD	YES	FALSE
19	765434	US	389	270	USD	NO	FALSE
20	123443	US	430	858	USD	YES	FALSE
21	674323	US	337	474	USD	YES	FALSE
22	323245	US	379	860	USD	NO	FALSE
23	765434	US	776	632	USD	YES	FALSE
24	674323	US	558	657	USD	YES	FALSE
25	674323	US	223	699	USD	NO	FALSE
26	674323	US	638	451	USD	YES	FALSE
27	123443	US	587	565	USD	YES	FALSE
28	674323	US	852	621	USD	NO	FALSE
29	323245	US	946	849	USD	YES	FALSE
30	323245	US	403	273	USD	YES	FALSE
31	123456	US	899	544	USD	NO	FALSE
32	123443	US	299	488	USD	YES	FALSE
33	674323	US	693	216	USD	YES	FALSE
34	123456	US	730	656	USD	NO	FALSE
35	765434	US	252	534	USD	YES	FALSE
36	674323	US	237	735	USD	YES	FALSE
38	674323	US	941	499	USD	YES	FALSE
39	323245	US	725	873	USD	YES	FALSE
40	122345	US	586	341	USD	NO	FALSE
41	545484	US	575	620	USD	YES	FALSE

5. As the last step a reconciliation report is to be generated from input file and aggregating the Error Report data.

tAggregateRow_1

Basic settings

Advanced settings

Dynamic settings

View

Documentation

Schema

Built-In

Edit schema

Sync columns

Group by

Output column

Input column position

Operations

Output column	Function	Input column position	Ignore null values
Error_Count	sum	Error_Count	<input type="checkbox"/>

Report				
Country	Total_Errors	Available_Products	Processed_Products	Rejected_Products
IN	12	150	140	10

Report				
Country	Total_Errors	Available_Products	Processed_Products	Rejected_Products
US	10	150	142	8

6. Once done with all the above steps, read the country wise processed data available.

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run

Kill

Clear

Product_ID	Country	Business_Rule	Attribute	Reason
2	IN	A	Product ID	Duplicate Product ID

tLogRow_1								
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery	OUT_OF_STOCK	errorMessage
5	674323	IN	19	null	INR	YES	FALSE	Price_Per_QTY!=null failed
8	null	IN	56	477	INR	YES	FALSE	Model!=null failed
13	765434	IN	707	647	null	NO	FALSE	Currency!=null failed
15	765434	null	590	473	INR	YES	FALSE	Country!=null failed
18	765434	null	575	425	INR	YES	FALSE	Country!=null failed
21	674323	IN	76	null	INR	YES	FALSE	Price_Per_QTY!=null failed
31	null	IN	167	438	INR	NO	FALSE	Model!=null failed
43	323245	IN	558	544	null	NO	FALSE	Currency!=null failed

Error Report					Error_Count
Product_ID	Country	Business_Rule	Attribute	Reason	Error_Count
null	IN	B	ID Price Currency	ID is null Price is null Currency is null	3
5	IN	B	Price	Price is null	1
8	IN	B	Model	Model is null	1
13	IN	B	Currency	Currency is null	1
15	null	B	Country	Country is null	1
18	null	B	Country	Country is null	1
21	IN	B	Price	Price is null	1
31	IN	B	Model	Model is null	1
43	IN	B	Currency	Currency is null	1

Report				
Country	Total_Errors	Available_Products	Processed_Products	Rejected_Products
IN	12	150	140	10

Basic Run

Debug Run

Advanced settings

Target Exec

Memory Run

Execution

Run

Kill

Clear

Duplicate_Product_ID				
Product_ID	Country	Business_Rule	Attribute	Reason
36	US	A	Product ID	Duplicate Product ID

tLogRow_2								
Product_ID	Model	Country	Available_QTY	Price_Per_QTY	Currency	Global_Delivery	OUT_OF_STOCK	errorMessage
3	null	US	12	494	USD	YES	FALSE	Model!=null failed
14	122345	US	134	null	USD	YES	FALSE	Price_Per_QTY!=null failed
15	765434	null	300	null	null	YES	FALSE	Country!=null failed Price_Per_QTY!=null failed Currency!=null failed
54	765434	US	794	null	USD	YES	FALSE	Price_Per_QTY!=null failed
117	765434	US	485	942	null	YES	FALSE	Currency!=null failed
141	674323	US	577	null	USD	YES	FALSE	Price_Per_QTY!=null failed
147	323245	US	743	547	null	YES	FALSE	Currency!=null failed

Error Report					Error_Count
Product_ID	Country	Business_Rule	Attribute	Reason	Error_Count
3	US	B	Model	Model is null	1
14	US	B	Price	Price is null	1
15	null	B	Country Price Currency	Country is null Price is null Currency is null	3
54	US	B	Price	Price is null	1
117	US	B	Currency	Currency is null	1
141	US	B	Price	Price is null	1
147	US	B	Currency	Currency is null	1

Report				
Country	Total_Errors	Available_Products	Processed_Products	Rejected_Products
US	10	150	142	8

