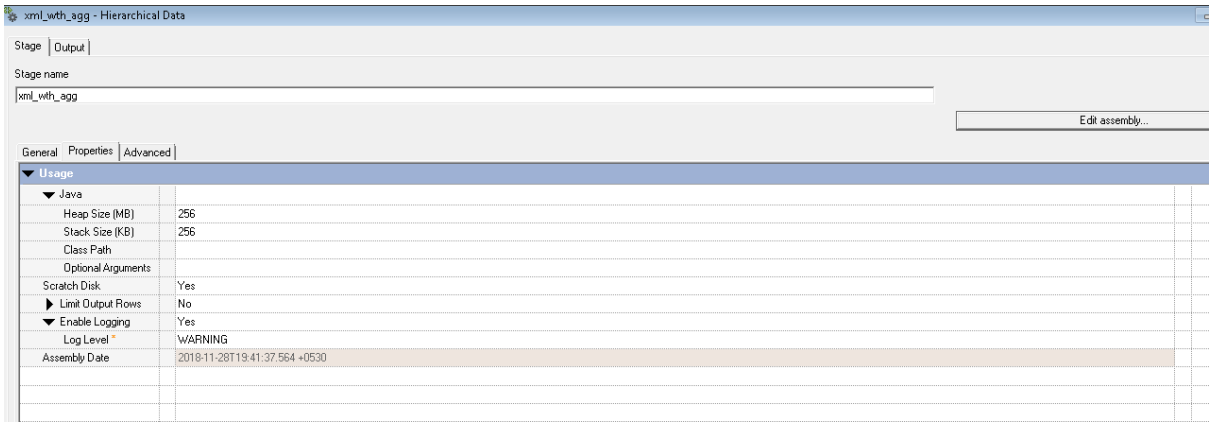
Hierarchical Stage in DataStage

Fig: Icon of the hierarchical stage.

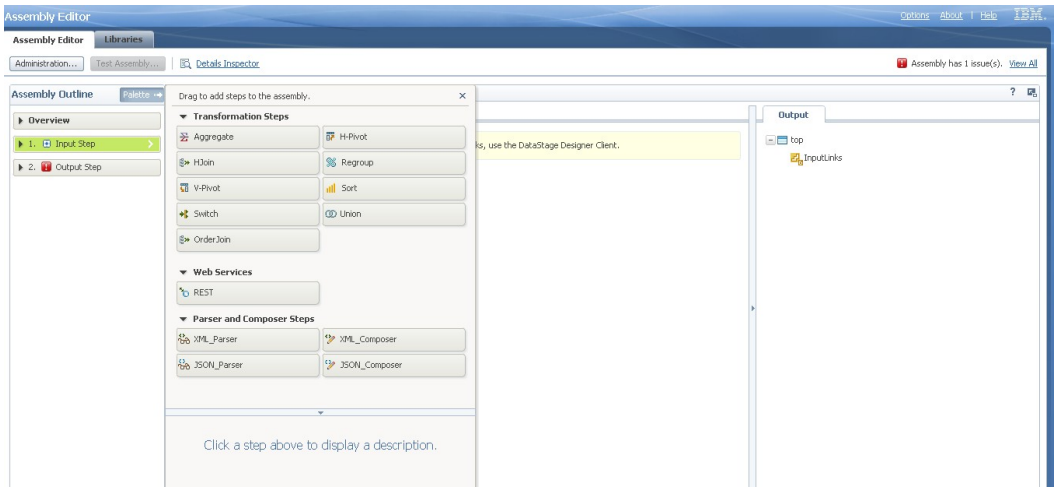
Hierarchical stage in DataStage is used to parse or compose XML (Extensible Markup Language) and JSON data. This stage was introduced in Version 11.3. When we have huge amounts of data to work with, then Hierarchical stage is preferred over XML packs.

Usage and benefits of Hierarchical stage:

- Extensible Markup Language (XML) data can be read in two ways:
  1. Using XML packs
  2. Using hierarchical data
- XML pack includes XML input, XML output, XML transformer stages, which are used for small transformations.
- Hierarchical data is used for complex transformations for large amounts of data.
- If we are working with large datasets then Hierarchical stage best choice always.
- Hierarchical data stage is used to create, parse and transform XML or JSON data.
- This stage is available in Realtime section of DataStage palette.
- In this article, we will explain about how XML data and JSON data is transformed.



- Above screenshot represents the homepage of the hierarchical stage.
- Click on edit assembly, below page with all the stages will appear from palette.
- Input step and output step are default steps.



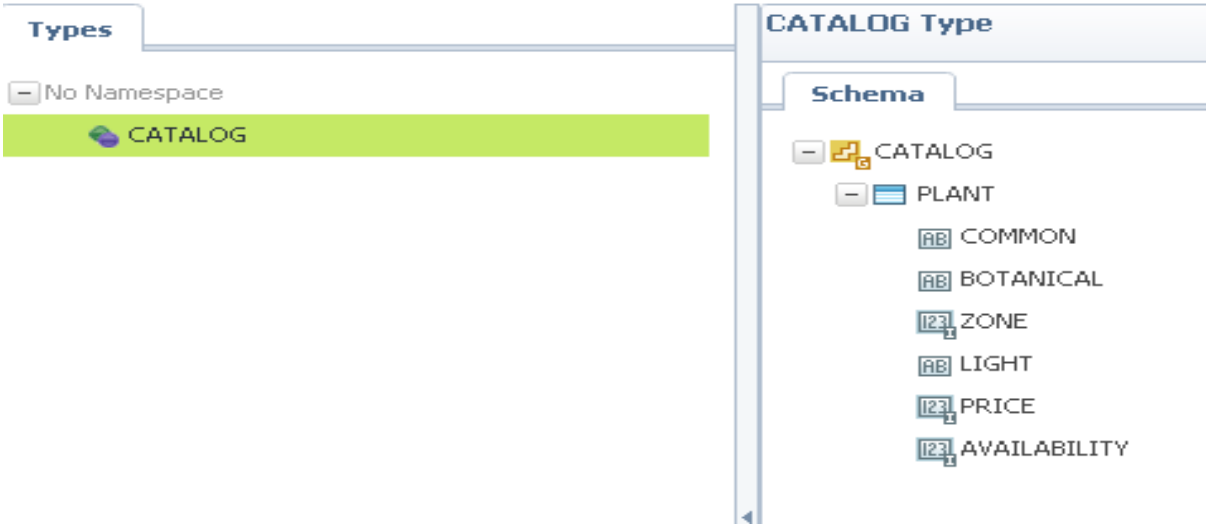
Example 1 - With aggregator step

A Job is created with an XML file with aggregation step.

Step 1:

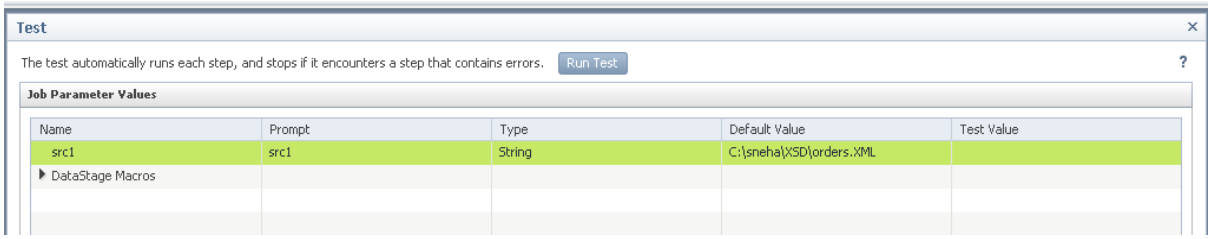
First of all we need to add the schema files in the library as below or from import table definitions.

- From libraries tab → click new library → give the library name, description and category.
- After creating library, click on the created library → click import new resource → add the schema files.
- Now we are able to see the columns.



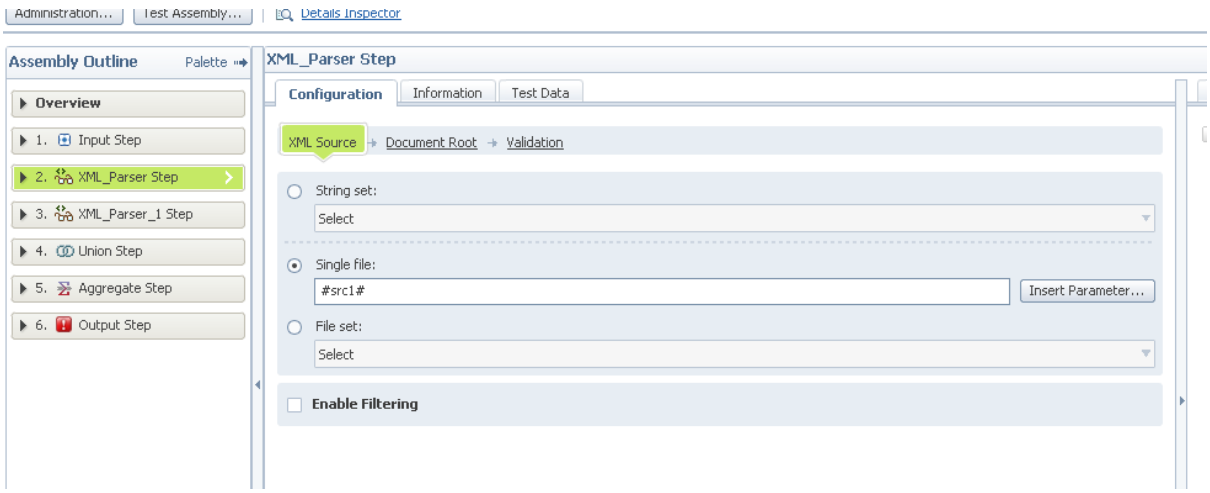
Step 2:

- Test the source file by clicking the test assembly tab → select the source file and click on run test.
- It will display as test completed if there are no errors.

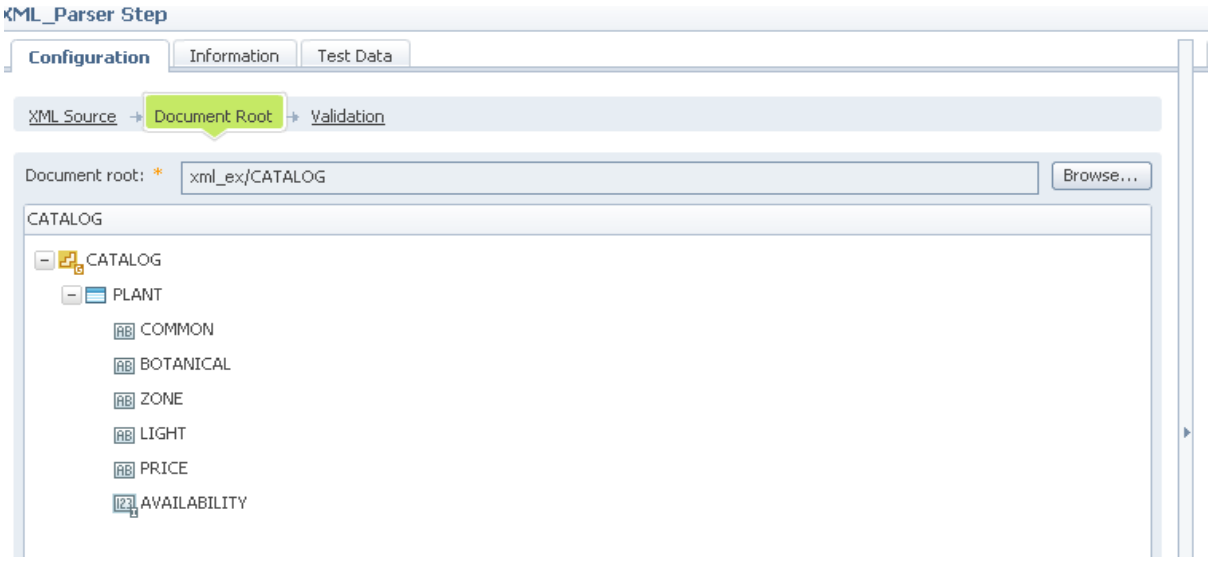


Step 3:

- Double click the XML-parser step from palette to take as input, which will appear after input step in assembly outline tab.
- XML Source tab contains String set, single file and file set.
- Check single file and click insert parameter →to take the source file (which appear from the parameters which we have given).



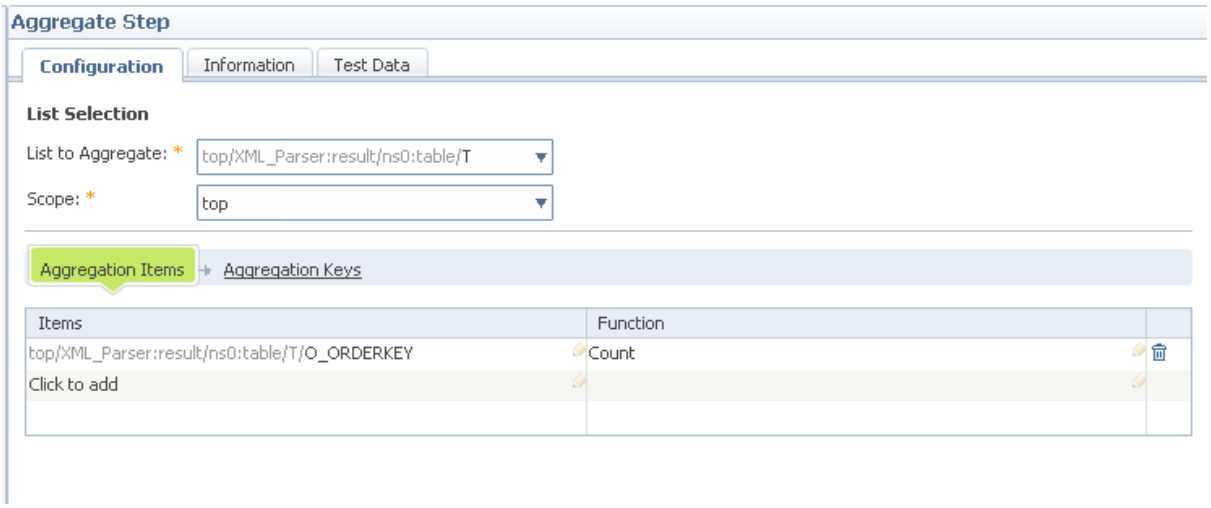
- In document root, click browse and get the XML file.

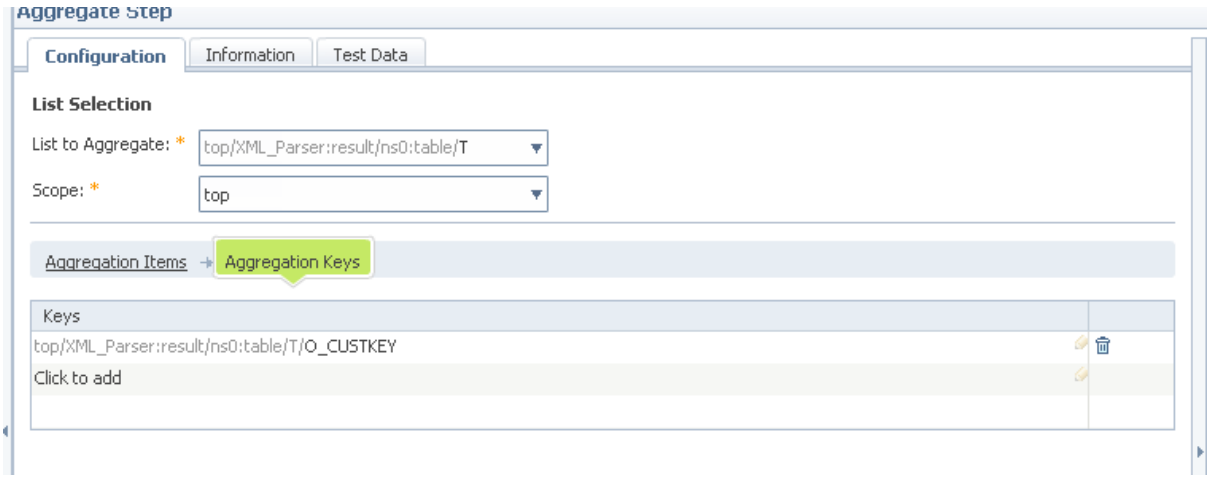


Step 4:

Double click on aggregate step from palette. This contains:

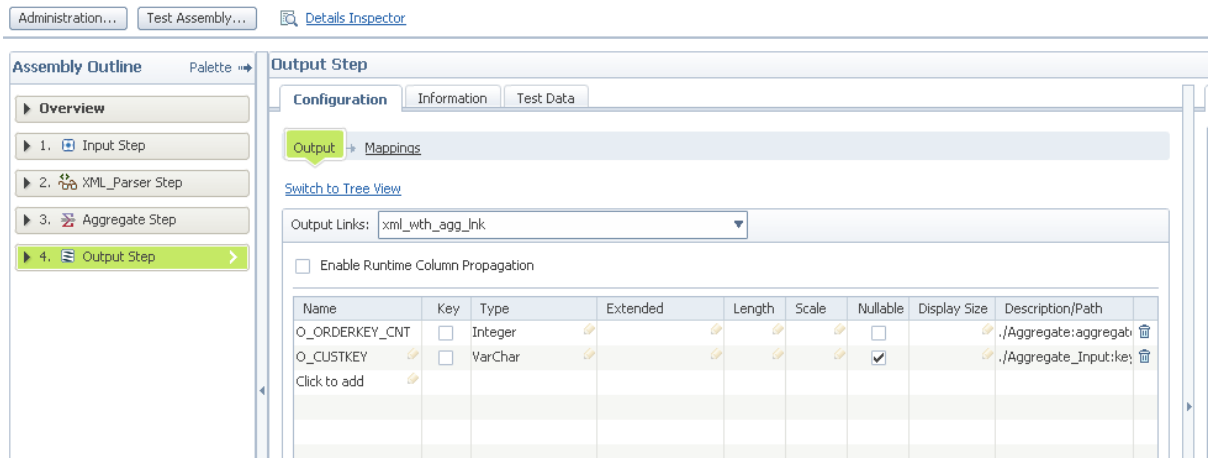
- **List to Aggregate:** Select the table from dropdown box.
- **Scope:** Select scope from dropdown.
- **Aggregation Items**: Select column for aggregation and type of function like sum, avg, count, min, max, concat.
- **Aggregation Keys**: Select group by column.





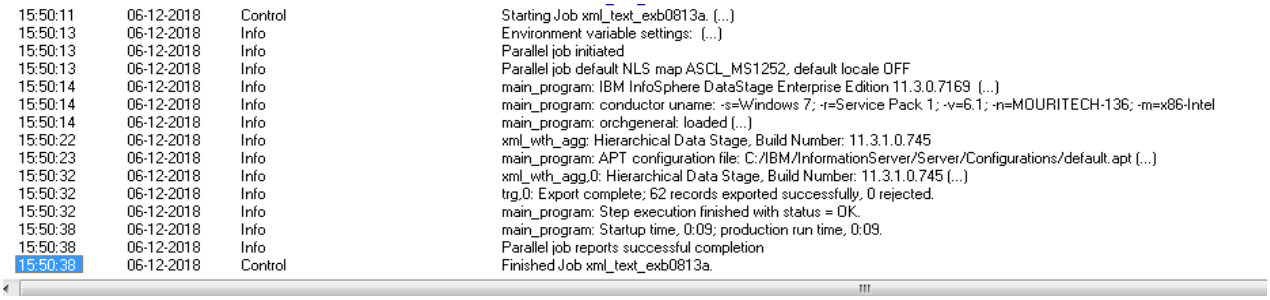
Step 5:

- This step contains columns to be displayed.
- Manually add the columns along with the datatype.
- In Mappings → map the columns for the given output columns.



Step 6:

- Click ok, Compile and run the job.
- Below screen is the log for the job:



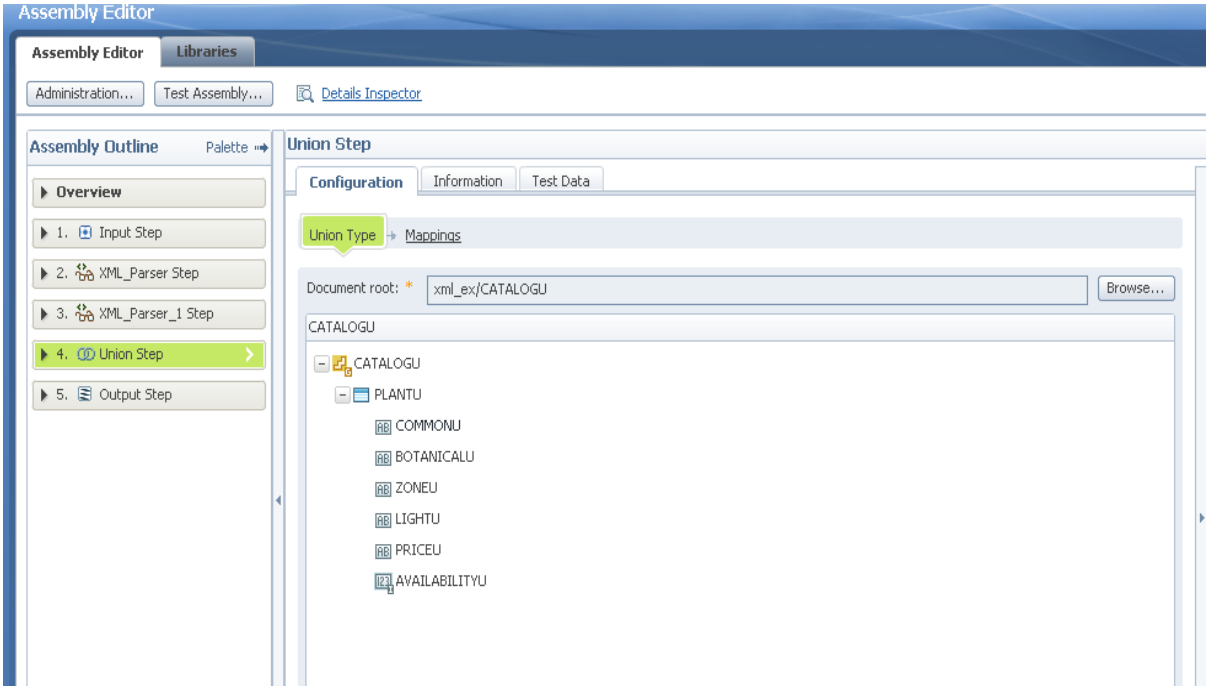
15:50:11	06-12-2018	Control	Starting Job xml_text_exb0813a. (...)
15:50:13	06-12-2018	Info	Environment variable settings: (...)
15:50:13	06-12-2018	Info	Parallel job initiated
15:50:13	06-12-2018	Info	Parallel job default NLS map ASCL_MS1252, default locale OFF
15:50:14	06-12-2018	Info	main_program: IBM InfoSphere DataStage Enterprise Edition 11.3.0.7169 (...)
15:50:14	06-12-2018	Info	main_program: conductor uname: -s=Windows 7; -r=Service Pack 1; -v=6.1; -n=MOURITECH-136; -m=x86-Intel
15:50:14	06-12-2018	Info	main_program: orchgeneral: loaded (...)
15:50:22	06-12-2018	Info	xml_wth_agg: Hierarchical Data Stage, Build Number: 11.3.1.0.745
15:50:23	06-12-2018	Info	main_program: APT configuration file: C:/IBM/InformationServer/Server/Configurations/default.apt (...)
15:50:32	06-12-2018	Info	xml_wth_agg,0: Hierarchical Data Stage, Build Number: 11.3.1.0.745 (...)
15:50:32	06-12-2018	Info	trg,0: Export complete; 62 records exported successfully, 0 rejected.
15:50:32	06-12-2018	Info	main_program: Step execution finished with status = OK.
15:50:38	06-12-2018	Info	main_program: Startup time, 0:09; production run time, 0:09.
15:50:38	06-12-2018	Info	Parallel job reports successful completion
15:50:38	06-12-2018	Control	Finished Job xml_text_exb0813a.

Output:

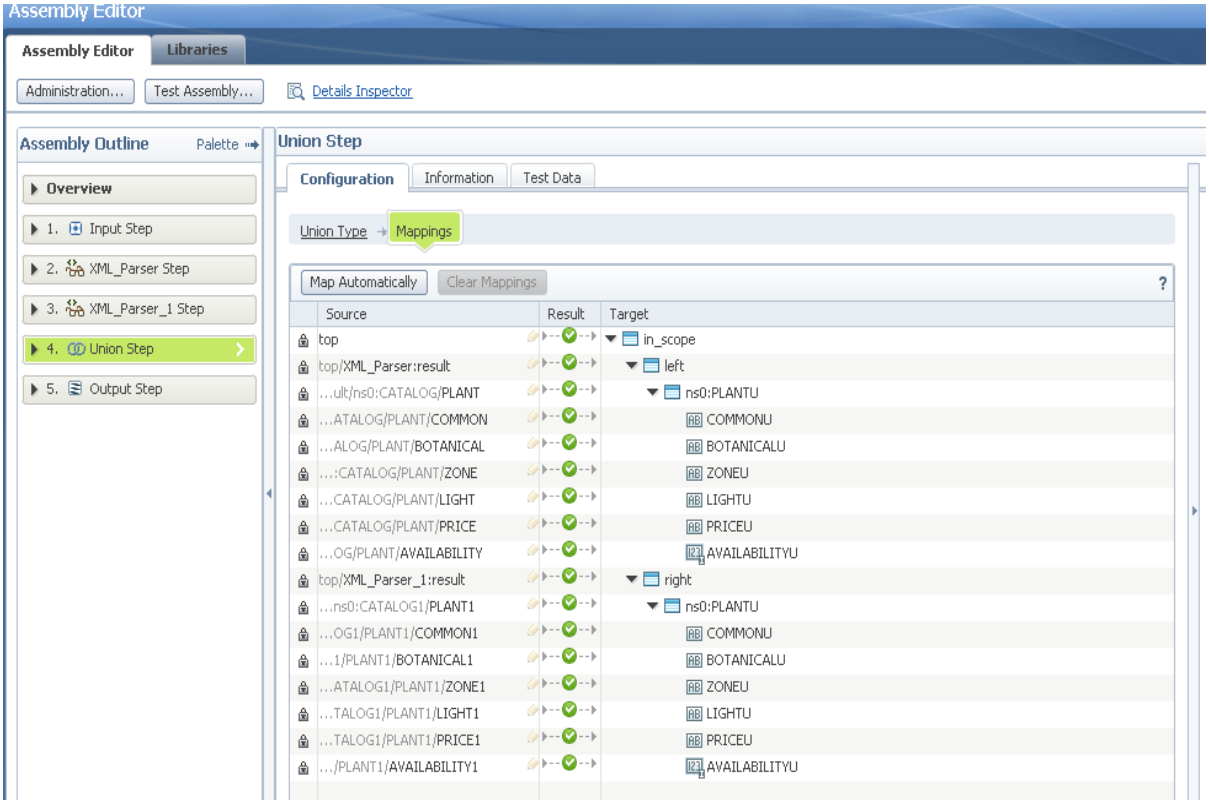
O_ORDERKEY_CNT	O_CUSTKEY
2	370
1	781
1	1234
1	1369
1	445
1	557
1	392
1	1301
1	670
1	611
2	1276
1	1153
1	862
2	1249
1	818
1	322
1	163
1	1292
1	568
1	286
1	845

Example 2 - With Union Step

- Add the input files.
- Double click on Union step and map the columns.



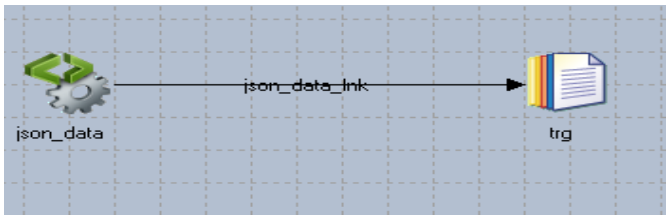
Here we are mapping XML parser, XML parser_1 metadata to output metadata.

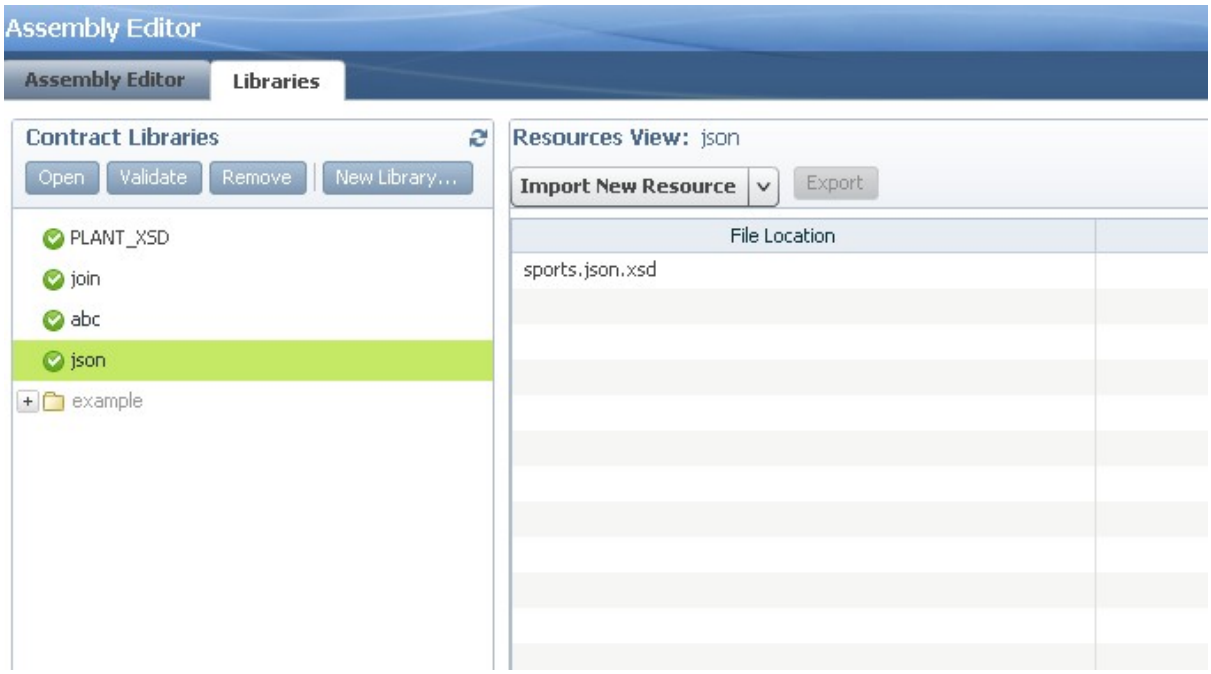


Output:

COMMONU	BOTANICALU	ZONEU	LIGHTU	PRICEU	AVAILABILITYU
Bloodroo	Sanguinari	4	Mostly S	2.44	31599
Columbin	Aquilegia	3	Mostly S	9.37	30699
Marsh Ma	Caltha pal	4	Mostly S	6.81	51799
Cowslip	Caltha pal	4	Mostly S	9.90	30699
Dutchman	Dicentra c	3	Mostly S	6.44	12099
Ginger,	Asarum can	3	Mostly S	9.03	41899
Hepatica	Hepatica a	4	Mostly S	4.45	12699
Liverlea	Hepatica a	4	Mostly S	3.99	10299
Jack-In-	Arisaema t	4	Mostly S	3.23	20199
Mayapple	Podophyllu	3	Mostly S	2.98	60599
Phlox, W	Phlox diva	3	Sun or S	2.80	12299
Phlox, B	Phlox diva	3	Sun or S	5.59	21699
Spring-B	Claytonia	7	Mostly S	6.59	20199
Trillium	Trillium g	5	Sun or S	3.90	42999
Wake Rob	Trillium g	5	Sun or S	3.20	22199
Violet,	Erythroniu	4	Shade	9.04	20199
Trout Li	Erythroniu	4	Shade	6.94	32499
Adder's-	Erythroniu	4	Shade	9.58	41399
Anemone	Anemone bl	6	Mostly S	8.86	122698
Grecian	Anemone bl	6	Mostly S	9.16	71099
Bee Balm	Monarda di	4	Shade	4.59	50399
Bergamot	Monarda di	4	Shade	7.16	42799
Black-Ey	Rudbeckia	Annual	Sunny	9.80	61899

Example 3 - Hierarchical stage with JSON data





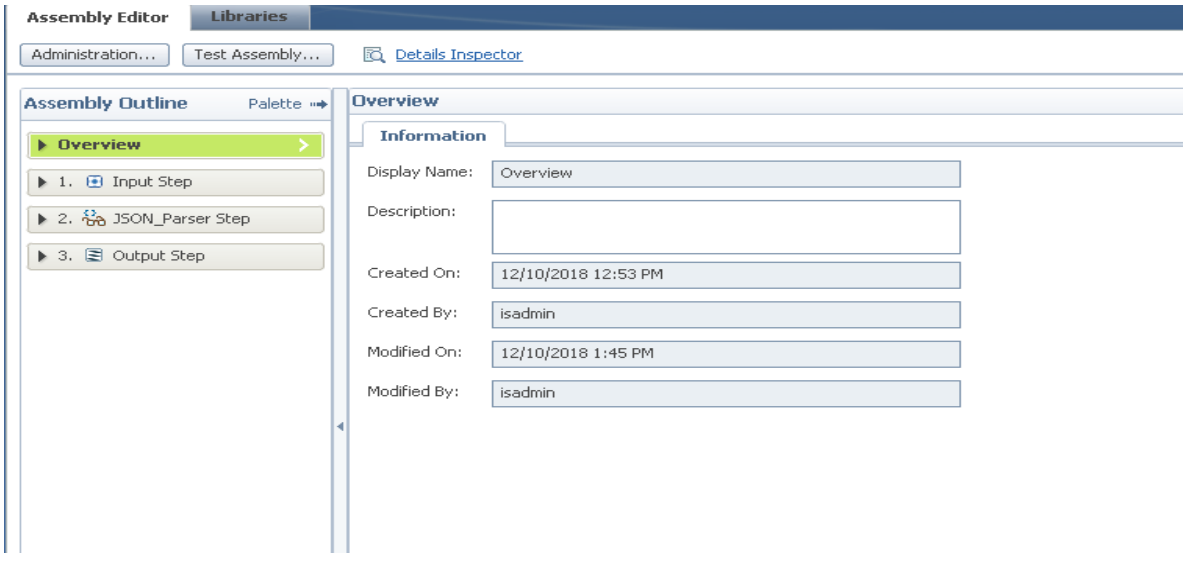
Step 1:

- First of all we need to import the XSD file.
- Go to Libraries → click new library → import new resource.

Step 2:

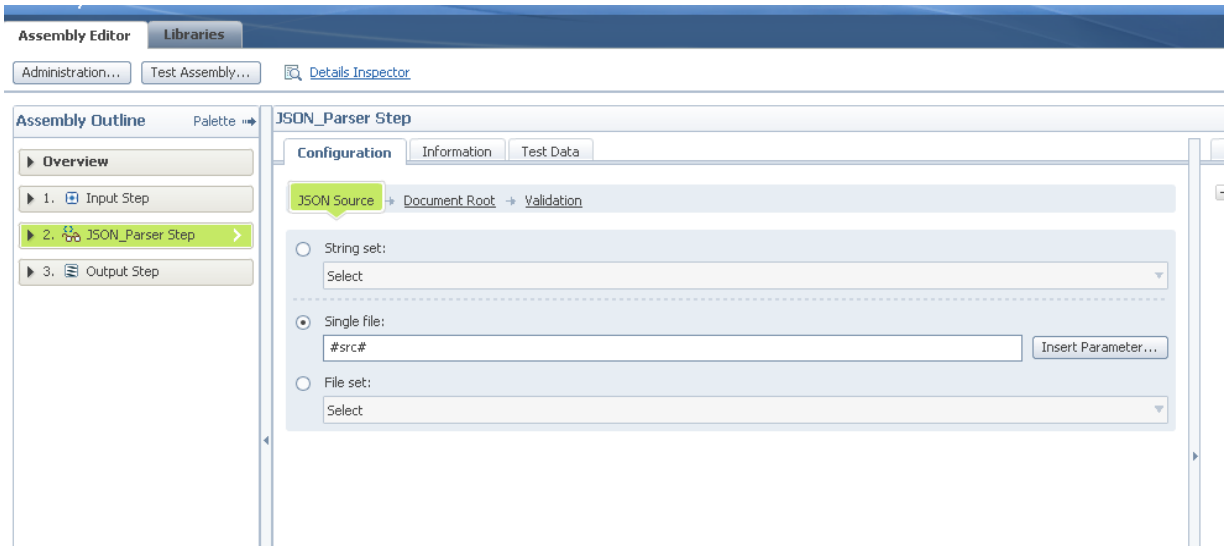
- Click assembly editor tab, by default input step and output step will be there.
- Double click on JSON_parser from palette to get in between input and output steps.

Source data:

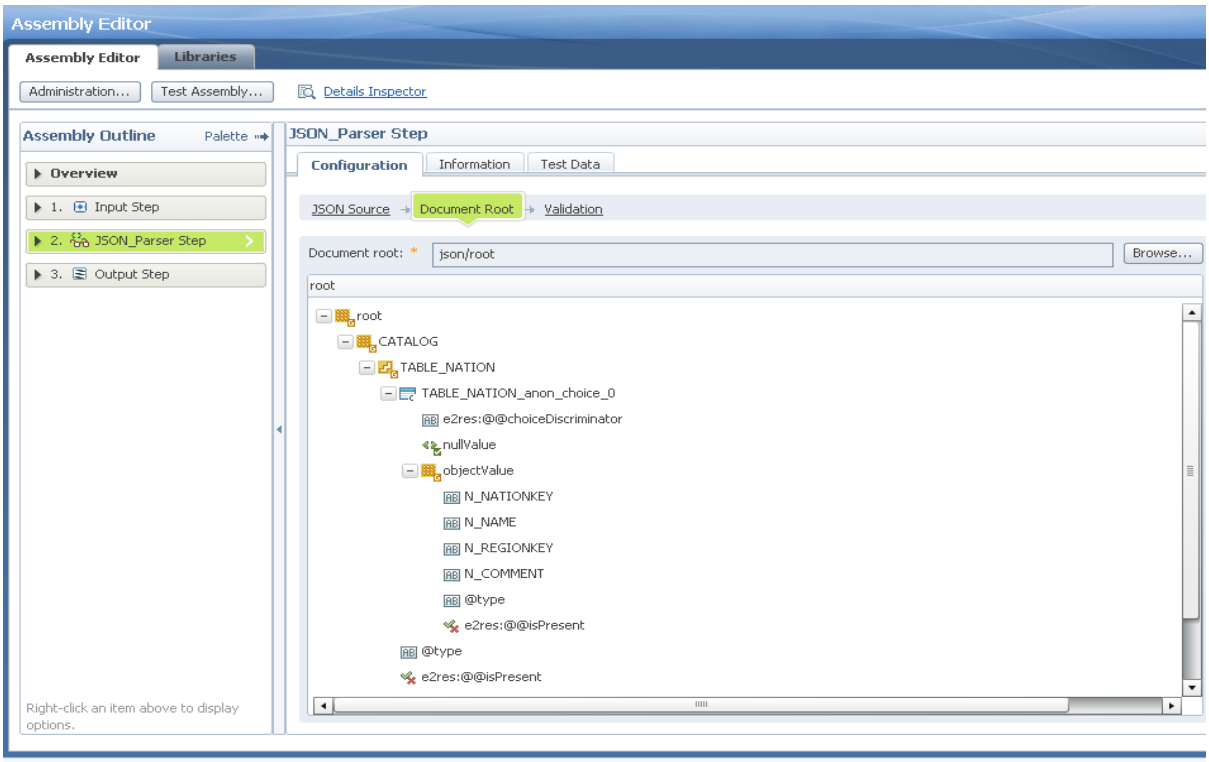


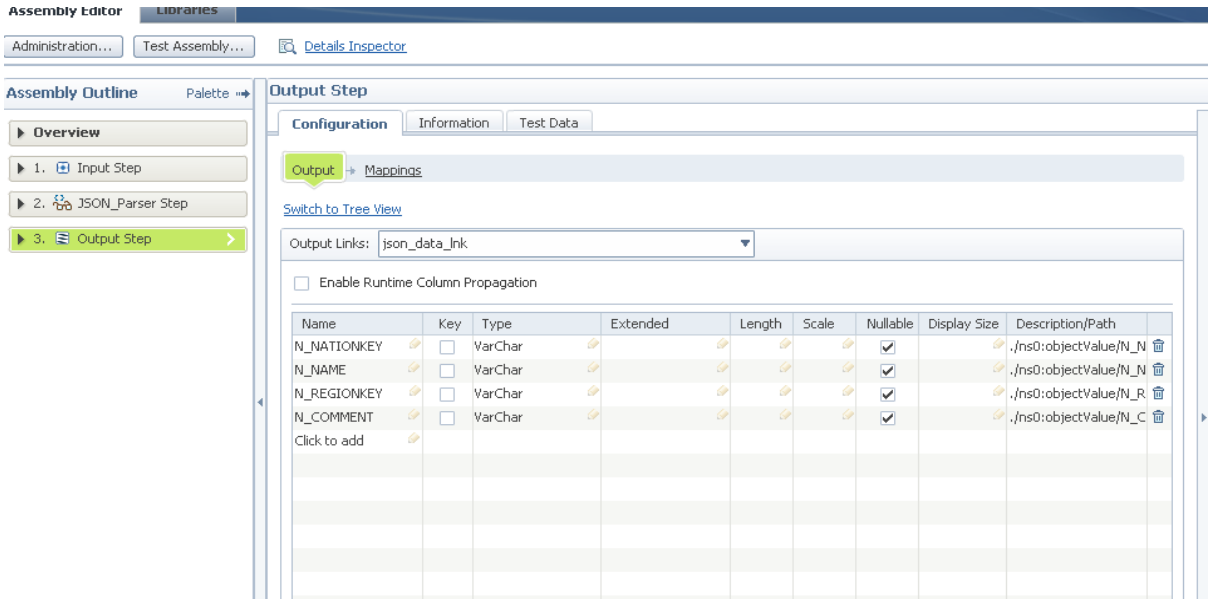
Step 3: In JSON_Parser step

- From JSON Source → click on single file radio button and insert the parameter of source file.

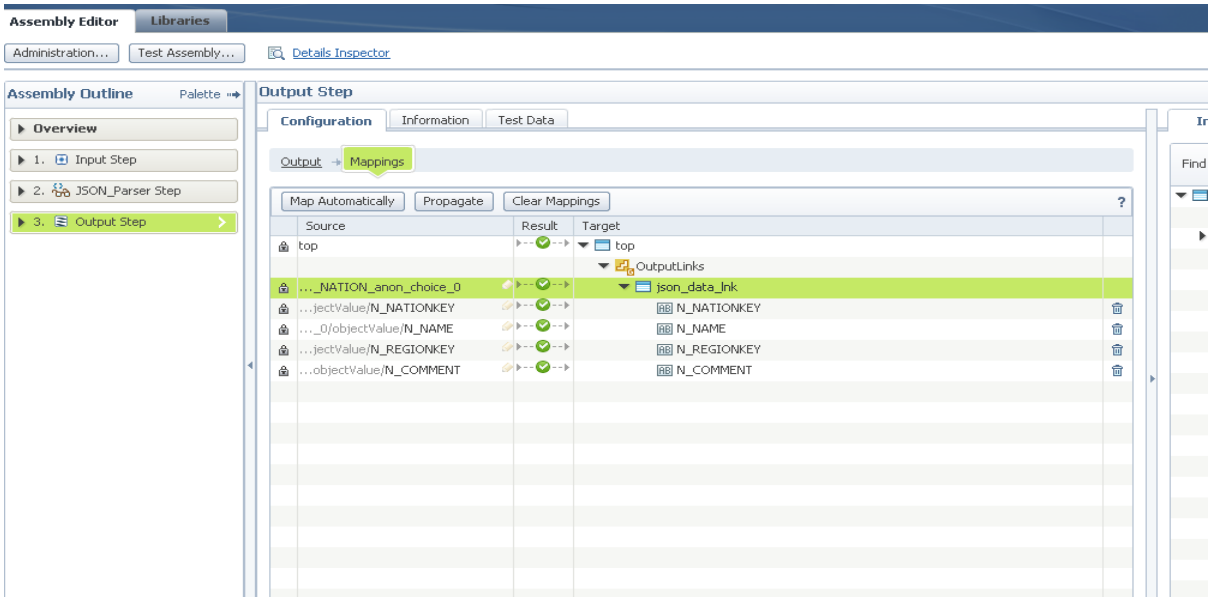


- In document root → click on browse to get the root file.





Click mappings and map the corresponding columns to the given output columns.



Step-5: Click ok and run the job.

- Below is the log of the job.

14:57:50	10-12-2018	Control	Starting Job Hirarchial_json. (...)
14:57:51	10-12-2018	Info	Environment variable settings: (...)
14:57:51	10-12-2018	Info	Parallel job initiated
14:57:51	10-12-2018	Info	Parallel job default NLS map ASCL_MS1252, default locale OFF
14:57:52	10-12-2018	Info	main_program: IBM InfoSphere DataStage Enterprise Edition 11.3.0.7169 (...)
14:57:52	10-12-2018	Info	main_program: conductor uname: -s=Windows 7; -r=Service Pack 1; -v=6.1; -n=MOURITECH-136; -m=x86-Intel
14:57:52	10-12-2018	Info	main_program: orchgeneral: loaded (...)
14:58:01	10-12-2018	Info	json_data: Hierarchical Data Stage, Build Number: 11.3.1.0.745
14:58:02	10-12-2018	Info	main_program: APT configuration file: C:/IBM/InformationServer/Server/Configurations/default.apt (...)
14:58:10	10-12-2018	Info	json_data,0: Hierarchical Data Stage, Build Number: 11.3.1.0.745 (...)
14:58:10	10-12-2018	Info	trg,0: Export complete; 24 records exported successfully, 0 rejected.
14:58:10	10-12-2018	Info	main_program: Step execution finished with status = OK.
14:58:16	10-12-2018	Info	main_program: Startup time, 0:08; production run time, 0:09.
14:58:16	10-12-2018	Info	Parallel job reports successful completion
14:58:16	10-12-2018	Control	Finished Job Hirarchial_json.

Output:

N_NATIONKEY	N_NAME	N_REGIONKEY	N_COMMENT
0	ALGERIA	0	final accounts wake q
1	ARGENTINA	1	idly final instructio
3	CANADA	1	foxes among the bold
4	EGYPT	4	pending accounts hagg
5	ETHIOPIA	0	fluffily ruthless req
6	FRANCE	3	even requests detect
7	GERMANY	3	blithely ironic foxes
8	INDIA	2	ironic packages shoul
9	INDONESIA	2	unusual excuses are q
10	IRAN	4	blithely even account
11	IRAQ	4	express, pending depo
12	JAPAN	2	blithely final packag
13	JORDAN	4	blithe, express depos
14	KENYA	0	ironic requests boost.
15	MOROCCO	0	ideas according to th
16	MOZAMBIQUE	0	ironic courts wake fl
17	PERU	1	final, final accounts
18	CHINA	2	bold accounts are. sl
19	ROMANIA	3	deposits boost against
20	SAUDI ARABIA	4	fluffily final account

Conclusion:

With the help of the above mentioned examples, it is useful for beginners who are working on Hierarchical stage instead of XML packs.

References:

www.ibm.com

https://www.ibm.com/support/knowledgecenter

CONTACT US FOR FURTHER DETAILS

Eswar Birukula

Team Lead, Analytics ETL Data Stage

eswarb.in@mouritech.com

MOURI Tech

www.mouritech.com