

Name: Shaikh Inamul Hasan

Roll No: 100

Assignment on input/output:

1. Retrieving and load Data from a dbms file (MS Access) to datawarehouse

The screenshot displays the Apache Spark IDE (Spoon) interface. The main workspace shows a transformation workflow with three components: 'Microsoft Excel input', 'Concat fields', and 'Microsoft Excel output', connected sequentially. The left sidebar contains a 'View' pane with a search bar and a list of output types, including 'Microsoft Excel output'. The bottom pane shows the 'Execution Results' tab, which includes a log of the transformation execution. The log indicates that the transformation started at 2023/10/21 20:23:51 and finished at 2023/10/21 20:23:52, with the following metrics: (I=5, O=0, R=0, W=5, U=0, E=0).

Execution Results

Logging Execution History Step Metrics Performance Graph Metrics Preview data

2023/10/21 20:23:51 - Spoon - Started the transformation execution.

2023/10/21 20:23:51 - Transformation 1 - Dispatching started for transformation [Transformation 1]

2023/10/21 20:23:52 - Microsoft Excel input.0 - Finished processing (I=5, O=0, R=0, W=5, U=0, E=0)

2023/10/21 20:23:52 - Concat fields.0 - Finished processing (I=0, O=0, R=5, W=5, U=0, E=0)

2023/10/21 20:23:52 - Microsoft Excel output.0 - Finished processing (I=0, O=5, R=5, W=5, U=0, E=0)

2023/10/21 20:23:52 - Spoon - The transformation has finished!!

2. Retrieving and load Data from a Flat File (Excel) to datawarehouse

Microsoft Excel inputNumber range → Table output

Examine preview data

Rows of step: Table output (5 rows)

#	ID	First Name	Last Name	Street	City	State	ZipCode	Age	Age Range
1	1.0	Mary	Joe	Thakur Marg	Mumbai	Maharastra	400101.0	5.0	Less than 5
2	2.0	Ram	Singh	LT Marg	Baroda	Gujrat	400206.0	15.0	10-15
3	3.0	Akshay	Kumar	SV Road	Lucknow	UP	400207.0	25.0	20-25
4	4.0	Viraj	Gupta	Linking Road	Jaipur	Rajasthan	400203.0	11.0	10-15
5	5.0	Samay	Khurana	WE highway	Nagpur	Maharastra	400102.0	18.0	15-20

Close

Logging Execution History Step Metrics Performance Graph Metrics Preview data

3. Retrieving and load Data from a RDBMS (MS SQL Server / ORACLE) to data warehouse (If you have oracle installed at home)

Table input → Split fields → Table output

Examine preview data

Rows of step: Table output (5 rows)

#	id	First Name	Last Name
1	11	Viraj	Gupta
2	12	Samay	Khurana
3	13	Ram	Singh
4	14	Mary	Joe
5	15	Akshay	Kumar

Close

Assignment on transformation:

1. Retrieve data from file staff.accdb and concat two columns firstname and lastname as fullname of teacher table and load data to data warehouse

Examine preview data

Rows of step: Table output (5 rows)

#	ID	First Name	Last Name	Street	City	State	FullName	ZipCode	Age	sal	Bonus	Full Name
1	1	Mary	Joe	Thakur Marg	Mumbai	Maharastra	<null>	400066	5	30000	1200	Mary Joe
2	2	Ram	Singh	LT Marg	Baroda	<null>	<null>	390001	15	35000	2300	Ram Singh
3	3	Akshay	Kumar	SV Road	Lucknow	UP	<null>	226001	25	50000	4500	Akshay Kumar
4	4	Viraj	Gupta	Linking Road	Jaipur	Rajasthan	<null>	302003	11	45000	3400	Viraj Gupta
5	5	Samay	Khurana	WE highway	Nagpur	Maharastra	<null>	440001	18	40000	3200	Samay Khurana

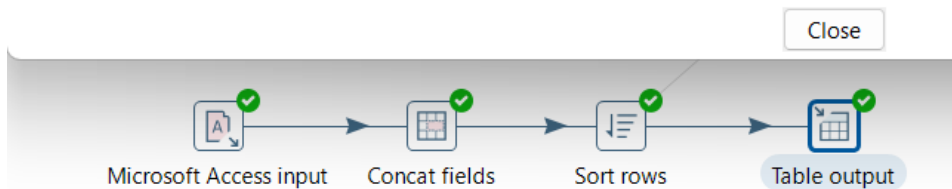
Close

Microsoft Access input → Concat fields → Table output

2. Sort column fullname in descending order and load data to datawarehouse.

Rows of step: Table output (5 rows)

#	ID	First Name	Last Name	fullname
1	4	Viraj	Gupta	Viraj Gupta
2	5	Samay	Khurana	Samay Khurana
3	2	Ram	Singh	Ram Singh
4	1	Mary	Joe	Mary Joe
5	3	Akshay	Kumar	Akshay Kumar



3. Retrieve data from file and descrtize age column of class table with > 5, 5-10, 10-15, 15-20, 20-25, >25 and load data to data warehouse

Microsoft Excel input → Table output

Examine preview data

Rows of step: Table output (5 rows)

#	ID	First Name	Last Name	Street	City	State	ZipCode	Age	Age Range
1	1.0	Mary	Joe	Thakur Marg	Mumbai	Maharastra	400101.0	5.0	Less than 5
2	2.0	Ram	Singh	LT Marg	Baroda	Gujrat	400206.0	15.0	10-15
3	3.0	Akshay	Kumar	SV Road	Lucknow	UP	400207.0	25.0	20-25
4	4.0	Viraj	Gupta	Linking Road	Jaipur	Rajasthan	400203.0	11.0	10-15
5	5.0	Samay	Khurana	WE highway	Nagpur	Maharastra	400102.0	18.0	15-20

Close

Logging Execution History Step Metrics Performance Graph Metrics Preview data

4. Retrieve data from file and add sequence in the staff table.

Microsoft Excel input → Add sequence → Table output

Examine preview data

Rows of step: Table output (5 rows)

#	ID	First Name	Last Name	Street	City	State	ZipCode	Age	Squence
1	1.0	Mary	Joe	Thakur Marg	Mumbai	Maharastra	400101.0	5.0	1
2	2.0	Ram	Singh	LT Marg	Baroda	Gujrat	400206.0	15.0	2
3	3.0	Akshay	Kumar	SV Road	Lucknow	UP	400207.0	25.0	3
4	4.0	Viraj	Gupta	Linking Road	Jaipur	Rajasthan	400203.0	11.0	4
5	5.0	Samay	Khurana	WE highway	Nagpur	Maharastra	400102.0	18.0	5

Close

5. Split fullname column into two columns Fname and Lname using staff table

Table input → Split fields → Table output

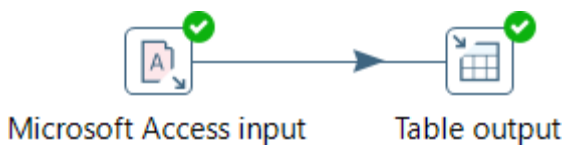
Examine preview data

Rows of step: Table output (5 rows)

#	id	First Name	Last Name
1	11	Viraj	Gupta
2	12	Samay	Khurana
3	13	Ram	Singh
4	14	Mary	Joe
5	15	Akshay	Kumar

Close

6. Retrieve data from file staff.accdb and set value of zip column of staff table in staffid column



ID	First Name	Last Name	Street	City	State	ZipCode	Age	Age Range
400066	Mary	Joe	Thakur Marg	Mumbai	Maharastra	NULL	5	NULL
390001	Ram	Singh	LT Marg	Baroda	NULL	NULL	15	NULL
226001	Akshay	Kumar	SV Road	Lucknow	UP	NULL	25	NULL
302003	Viraj	Gupta	Linking Road	Jaipur	Rajasthan	NULL	11	NULL
440001	Samay	Khurana	WE highway	Nagpur	Maharastra	NULL	18	NULL

5 rows in set (0.00 sec)

7. Load data transformed in step 1-3 in another MS – Accessdatabase.

The diagram shows a sequence of steps: 'Microsoft Access input' (document icon) → 'Concat fields' (grid icon) → 'Sort rows' (list icon) → 'Number range' (list icon) → 'Table output' (grid icon). All icons have green checkmarks.

Below the diagram is a screenshot of the 'Examine preview data' window. It displays the following data:

#	ID	First Name	Last Name	Street	City	State	FullName	ZipCode	Age	sal	Bonus	Full Name	Age Range
1	1	Mary	Joe	Thakur Marg	Mumbai	Maharastra	<null>	400066	5	30000	1200	Mary Joe	Age Range
2	2	Ram	Singh	LT Marg	Baroda	<null>	<null>	390001	15	35000	2300	Ram Singh	Age Range
3	3	Akshay	Kumar	SV Road	Lucknow	UP	<null>	226001	25	50000	4500	Akshay Kumar	Age Range
4	4	Viraj	Gupta	Linking Road	Jaipur	Rajasthan	<null>	302003	11	45000	3400	Viraj Gupta	Age Range
5	5	Samay	Khurana	WE highway	Nagpur	Maharastra	<null>	440001	18	40000	3200	Samay Khurana	Age Range

Close

2023/10/21 01:35:38 - Number range:0 - Finished processing (I=0, O=0, R=3, W=3, U=0, E=0)

8. Retrieve data from file staff.accdb and apply following stringoperation

a. Convert city column in upper case

b. Convert state column in lower case

c. Apply left padding in city and fix width 20 and padding characters should be “*”

Examine preview data

Rows of step: Table output (5 rows)

#	ID	First Name	Last Name	Street	City	State	FullName	ZipCode	Age	sal	Bonus
1	1	Mary	Joe	Thakur Marg	*****MUMBAI	maharashtra	<null>	400066	5	30000	1200
2	2	Ram	Singh	LT Marg	*****BARODA	<null>	<null>	390001	15	35000	2300
3	3	Akshay	Kumar	SV Road	*****LUCKNOW	up	<null>	226001	25	50000	4500
4	4	Viraj	Gupta	Linking Road	*****JAIPUR	rajasthan	<null>	302003	11	45000	3400
5	5	Samay	Khurana	WE highway	*****NAGPUR	maharashtra	<null>	440001	18	40000	3200

Close

Microsoft Access input String operations Table output

9. Retrieve data from file staff.accdb and calculate total salary using which is sum of salary and bonus column

Microsoft Access input Calculator

Examine preview data

Rows of step: Text file output (5 rows)

#	ID	First Name	Last Name	Street	City	State	FullName	ZipCode	Age	sal	Bonus	Total Salary
1	1	Mary	Joe	Thakur Marg	Mumbai	Maharashtra	<null>	400066	5	30000	1200	31200
2	2	Ram	Singh	LT Marg	Baroda	<null>	<null>	390001	15	35000	2300	37300
3	3	Akshay	Kumar	SV Road	Lucknow	UP	<null>	226001	25	50000	4500	54500
4	4	Viraj	Gupta	Linking Road	Jaipur	Rajasthan	<null>	302003	11	45000	3400	48400
5	5	Samay	Khurana	WE highway	Nagpur	Maharashtra	<null>	440001	18	40000	3200	43200

Close