

Assignment #2 - Data Transform Activity

1-a) Creating a new table for my analysis called "employee_sales"

Services

- HDFS
- YARN
- MapReduce2
- Tez
- Hive
- Pig
- Sqoop
- Oozie
- ZooKeeper
- Ambari Metrics
- WebHCat
- Hosts

DATABASE

Select or search database/schema

default

Browse

```

1 create table employee_sales(
2   Age INT,
3   Attrition STRING,
4   BusinessTravel STRING,
5   DailyRate INT,
6   Department STRING,
7   DistanceFromHome INT,
8   Education STRING,
9   EducationField INT,
10  EmployeeCount INT,
11  EmployeeNumber INT,
12  EnvironmentSatisfaction INT,
13  Gender STRING,
14  HourlyRate INT,
15  JobTitle STRING,
16  MaritalStatus STRING,
17  MonthlyIncome INT,
18  NumCompaniesWorked INT,
19  OverTime STRING,
20  ParterRate INT,
21  PercentHike INT,
22  PercentLeave INT,
23  PercentPromoted INT,
24  PercentStockOption INT,
25  PercentTravel INT,
26  Salary INT,
27  StandardHours INT,
28  TotalCompensation INT,
29  TotalPay INT,
30  TotalPayPerHour INT,
31  TotalPayPerMonth INT,
32  TotalPayPerYear INT,
33  TotalPayPerYear2 INT,
34  TotalPayPerYear3 INT,
35  TotalPayPerYear4 INT,
36  TotalPayPerYear5 INT,
37  TotalPayPerYear6 INT,
38  TotalPayPerYear7 INT,
39  TotalPayPerYear8 INT,
40  TotalPayPerYear9 INT,
41  TotalPayPerYear10 INT,
42  TotalPayPerYear11 INT,
43  TotalPayPerYear12 INT,
44  TotalPayPerYear13 INT,
45  TotalPayPerYear14 INT,
46  TotalPayPerYear15 INT,
47  TotalPayPerYear16 INT,
48  TotalPayPerYear17 INT,
49  TotalPayPerYear18 INT,
50  TotalPayPerYear19 INT,
51  TotalPayPerYear20 INT,
52  TotalPayPerYear21 INT,
53  TotalPayPerYear22 INT,
54  TotalPayPerYear23 INT,
55  TotalPayPerYear24 INT,
56  TotalPayPerYear25 INT,
57  TotalPayPerYear26 INT,
58  TotalPayPerYear27 INT,
59  TotalPayPerYear28 INT,
60  TotalPayPerYear29 INT,
61  TotalPayPerYear30 INT,
62  TotalPayPerYear31 INT,
63  TotalPayPerYear32 INT,
64  TotalPayPerYear33 INT,
65  TotalPayPerYear34 INT,
66  TotalPayPerYear35 INT,
67  TotalPayPerYear36 INT,
68  TotalPayPerYear37 INT,
69  TotalPayPerYear38 INT,
70  TotalPayPerYear39 INT,
71  TotalPayPerYear40 INT,
72  TotalPayPerYear41 INT,
73  TotalPayPerYear42 INT,
74  TotalPayPerYear43 INT,
75  TotalPayPerYear44 INT,
76  TotalPayPerYear45 INT,
77  TotalPayPerYear46 INT,
78  TotalPayPerYear47 INT,
79  TotalPayPerYear48 INT,
80  TotalPayPerYear49 INT,
81  TotalPayPerYear50 INT,
82  TotalPayPerYear51 INT,
83  TotalPayPerYear52 INT,
84  TotalPayPerYear53 INT,
85  TotalPayPerYear54 INT,
86  TotalPayPerYear55 INT,
87  TotalPayPerYear56 INT,
88  TotalPayPerYear57 INT,
89  TotalPayPerYear58 INT,
90  TotalPayPerYear59 INT,
91  TotalPayPerYear60 INT,
92  TotalPayPerYear61 INT,
93  TotalPayPerYear62 INT,
94  TotalPayPerYear63 INT,
95  TotalPayPerYear64 INT,
96  TotalPayPerYear65 INT,
97  TotalPayPerYear66 INT,
98  TotalPayPerYear67 INT,
99  TotalPayPerYear68 INT,
100 TotalPayPerYear69 INT,
101 TotalPayPerYear70 INT,
102 TotalPayPerYear71 INT,
103 TotalPayPerYear72 INT,
104 TotalPayPerYear73 INT,
105 TotalPayPerYear74 INT,
106 TotalPayPerYear75 INT,
107 TotalPayPerYear76 INT,
108 TotalPayPerYear77 INT,
109 TotalPayPerYear78 INT,
110 TotalPayPerYear79 INT,
111 TotalPayPerYear80 INT,
112 TotalPayPerYear81 INT,
113 TotalPayPerYear82 INT,
114 TotalPayPerYear83 INT,
115 TotalPayPerYear84 INT,
116 TotalPayPerYear85 INT,
117 TotalPayPerYear86 INT,
118 TotalPayPerYear87 INT,
119 TotalPayPerYear88 INT,
120 TotalPayPerYear89 INT,
121 TotalPayPerYear90 INT,
122 TotalPayPerYear91 INT,
123 TotalPayPerYear92 INT,
124 TotalPayPerYear93 INT,
125 TotalPayPerYear94 INT,
126 TotalPayPerYear95 INT,
127 TotalPayPerYear96 INT,
128 TotalPayPerYear97 INT,
129 TotalPayPerYear98 INT,
130 TotalPayPerYear99 INT,
131 TotalPayPerYear100 INT,
132 TotalPayPerYear101 INT,
133 TotalPayPerYear102 INT,
134 TotalPayPerYear103 INT,
135 TotalPayPerYear104 INT,
136 TotalPayPerYear105 INT,
137 TotalPayPerYear106 INT,
138 TotalPayPerYear107 INT,
139 TotalPayPerYear108 INT,
140 TotalPayPerYear109 INT,
141 TotalPayPerYear110 INT,
142 TotalPayPerYear111 INT,
143 TotalPayPerYear112 INT,
144 TotalPayPerYear113 INT,
145 TotalPayPerYear114 INT,
146 TotalPayPerYear115 INT,
147 TotalPayPerYear116 INT,
148 TotalPayPerYear117 INT,
149 TotalPayPerYear118 INT,
150 TotalPayPerYear119 INT,
151 TotalPayPerYear120 INT,
152 TotalPayPerYear121 INT,
153 TotalPayPerYear122 INT,
154 TotalPayPerYear123 INT,
155 TotalPayPerYear124 INT,
156 TotalPayPerYear125 INT,
157 TotalPayPerYear126 INT,
158 TotalPayPerYear127 INT,
159 TotalPayPerYear128 INT,
160 TotalPayPerYear129 INT,
161 TotalPayPerYear130 INT,
162 TotalPayPerYear131 INT,
163 TotalPayPerYear132 INT,
164 TotalPayPerYear133 INT,
165 TotalPayPerYear134 INT,
166 TotalPayPerYear135 INT,
167 TotalPayPerYear136 INT,
168 TotalPayPerYear137 INT,
169 TotalPayPerYear138 INT,
170 TotalPayPerYear139 INT,
171 TotalPayPerYear140 INT,
172 TotalPayPerYear141 INT,
173 TotalPayPerYear142 INT,
174 TotalPayPerYear143 INT,
175 TotalPayPerYear144 INT,
176 TotalPayPerYear145 INT,
177 TotalPayPerYear146 INT,
178 TotalPayPerYear147 INT,
179 TotalPayPerYear148 INT,
180 TotalPayPerYear149 INT,
181 TotalPayPerYear150 INT,
182 TotalPayPerYear151 INT,
183 TotalPayPerYear152 INT,
184 TotalPayPerYear153 INT,
185 TotalPayPerYear154 INT,
186 TotalPayPerYear155 INT,
187 TotalPayPerYear156 INT,
188 TotalPayPerYear157 INT,
189 TotalPayPerYear158 INT,
190 TotalPayPerYear159 INT,
191 TotalPayPerYear160 INT,
192 TotalPayPerYear161 INT,
193 TotalPayPerYear162 INT,
194 TotalPayPerYear163 INT,
195 TotalPayPerYear164 INT,
196 TotalPayPerYear165 INT,
197 TotalPayPerYear166 INT,
198 TotalPayPerYear167 INT,
199 TotalPayPerYear168 INT,
200 TotalPayPerYear169 INT,
201 TotalPayPerYear170 INT,
202 TotalPayPerYear171 INT,
203 TotalPayPerYear172 INT,
204 TotalPayPerYear173 INT,
205 TotalPayPerYear174 INT,
206 TotalPayPerYear175 INT,
207 TotalPayPerYear176 INT,
208 TotalPayPerYear177 INT,
209 TotalPayPerYear178 INT,
210 TotalPayPerYear179 INT,
211 TotalPayPerYear180 INT,
212 TotalPayPerYear181 INT,
213 TotalPayPerYear182 INT,
214 TotalPayPerYear183 INT,
215 TotalPayPerYear184 INT,
216 TotalPayPerYear185 INT,
217 TotalPayPerYear186 INT,
218 TotalPayPerYear187 INT,
219 TotalPayPerYear188 INT,
220 TotalPayPerYear189 INT,
221 TotalPayPerYear190 INT,
222 TotalPayPerYear191 INT,
223 TotalPayPerYear192 INT,
224 TotalPayPerYear193 INT,
225 TotalPayPerYear194 INT,
226 TotalPayPerYear195 INT,
227 TotalPayPerYear196 INT,
228 TotalPayPerYear197 INT,
229 TotalPayPerYear198 INT,
230 TotalPayPerYear199 INT,
231 TotalPayPerYear200 INT,
232 TotalPayPerYear201 INT,
233 TotalPayPerYear202 INT,
234 TotalPayPerYear203 INT,
235 TotalPayPerYear204 INT,
236 TotalPayPerYear205 INT,
237 TotalPayPerYear206 INT,
238 TotalPayPerYear207 INT,
239 TotalPayPerYear208 INT,
240 TotalPayPerYear209 INT,
241 TotalPayPerYear210 INT,
242 TotalPayPerYear211 INT,
243 TotalPayPerYear212 INT,
244 TotalPayPerYear213 INT,
245 TotalPayPerYear214 INT,
246 TotalPayPerYear215 INT,
247 TotalPayPerYear216 INT,
248 TotalPayPerYear217 INT,
249 TotalPayPerYear218 INT,
250 TotalPayPerYear219 INT,
251 TotalPayPerYear220 INT,
252 TotalPayPerYear221 INT,
253 TotalPayPerYear222 INT,
254 TotalPayPerYear223 INT,
255 TotalPayPerYear224 INT,
256 TotalPayPerYear225 INT,
257 TotalPayPerYear226 INT,
258 TotalPayPerYear227 INT,
259 TotalPayPerYear228 INT,
260 TotalPayPerYear229 INT,
261 TotalPayPerYear230 INT,
262 TotalPayPerYear231 INT,
263 TotalPayPerYear232 INT,
264 TotalPayPerYear233 INT,
265 TotalPayPerYear234 INT,
266 TotalPayPerYear235 INT,
267 TotalPayPerYear236 INT,
268 TotalPayPerYear237 INT,
269 TotalPayPerYear238 INT,
270 TotalPayPerYear239 INT,
271 TotalPayPerYear240 INT,
272 TotalPayPerYear241 INT,
273 TotalPayPerYear242 INT,
274 TotalPayPerYear243 INT,
275 TotalPayPerYear244 INT,
276 TotalPayPerYear245 INT,
277 TotalPayPerYear246 INT,
278 TotalPayPerYear247 INT,
279 TotalPayPerYear248 INT,
280 TotalPayPerYear249 INT,
281 TotalPayPerYear250 INT,
282 TotalPayPerYear251 INT,
283 TotalPayPerYear252 INT,
284 TotalPayPerYear253 INT,
285 TotalPayPerYear254 INT,
286 TotalPayPerYear255 INT,
287 TotalPayPerYear256 INT,
288 TotalPayPerYear257 INT,
289 TotalPayPerYear258 INT,
290 TotalPayPerYear259 INT,
291 TotalPayPerYear260 INT,
292 TotalPayPerYear261 INT,
293 TotalPayPerYear262 INT,
294 TotalPayPerYear263 INT,
295 TotalPayPerYear264 INT,
296 TotalPayPerYear265 INT,
297 TotalPayPerYear266 INT,
298 TotalPayPerYear267 INT,
299 TotalPayPerYear268 INT,
300 TotalPayPerYear269 INT,
301 TotalPayPerYear270 INT,
302 TotalPayPerYear271 INT,
303 TotalPayPerYear272 INT,
304 TotalPayPerYear273 INT,
305 TotalPayPerYear274 INT,
306 TotalPayPerYear275 INT,
307 TotalPayPerYear276 INT,
308 TotalPayPerYear277 INT,
309 TotalPayPerYear278 INT,
310 TotalPayPerYear279 INT,
311 TotalPayPerYear280 INT,
312 TotalPayPerYear281 INT,
313 TotalPayPerYear282 INT,
314 TotalPayPerYear283 INT,
315 TotalPayPerYear284 INT,
316 TotalPayPerYear285 INT,
317 TotalPayPerYear286 INT,
318 TotalPayPerYear287 INT,
319 TotalPayPerYear288 INT,
320 TotalPayPerYear289 INT,
321 TotalPay
```

b) Loading the table "employee" into this table.

The screenshot shows the Ambari Query Editor interface. The top navigation bar includes links for QUERY, JOBS, TABLES, SAVED QUERIES, UDFs, SETTINGS, and NOTIFICATIONS. The left sidebar contains a menu with options like Dashboard, Services, HDFS, YARN, MapReduce2, Tez, Hive, Pig, Sqoop, Oozie, and ZooKeeper. The main workspace is divided into several sections. At the top, there are tabs for Worksheet1, Worksheet3, and Worksheet4. Below the tabs, there is a 'DATABASE' dropdown menu currently set to 'default', with a 'Browse' button to its right. The central area is a text editor containing a SQL query:
1 insert overwrite table employee_sales;
2 select * from employee;
To the right of the text editor, there is a panel titled 'default Tables(3)' which lists the available tables: 'employee', 'employee_sales', and 'hivesampletable'.

Ambari

Dashboard

Services

- HDFS
- YARN
- MapReduce2
- Tez
- Hive
- Pig
- Sqoop
- Oozie

DATABASE

Select or search database/schema

default

Browse

```
1 select * from employee_sales;
```

default Tables(3)

Search Tables

- employee
- employee_sales
- hivesampletable

employee_sales.age employee_sales.attrition employee_sales.businesstravel employee_sales.dailyrate

41	Yes	Travel_Rarely	1102
49	No	Travel_Frequently	279
37	Yes	Travel_Rarely	1373
33	No	Travel_Frequently	1392
27	No	Travel_Rarely	591
32	No	Travel_Frequently	1005
59	No	Travel_Rarely	1324
30	No	Travel_Rarely	1358
38	No	Travel_Frequently	216
36	No	Travel_Rarely	1299
35	No	Travel_Rarely	809
29	No	Travel_Rarely	153
31	No	Travel_Rarely	670
34	No	Travel_Rarely	1346

Type here to search

5°C

5:09 PM 2021-11-02

c) Selecting these columns: Attrition, Department, Job Satisfaction & Monthly Income

Ambari

Dashboard

Services

- HDFS
- YARN
- MapReduce2
- Tez
- Hive
- Pig
- Sqoop
- Oozie
- ZooKeeper
- Ambari Metrics

QUERY

Worksheet1 * Worksheet3 * Worksheet4 * Worksheet5 *

DATABASE

Select or search database/schema

default

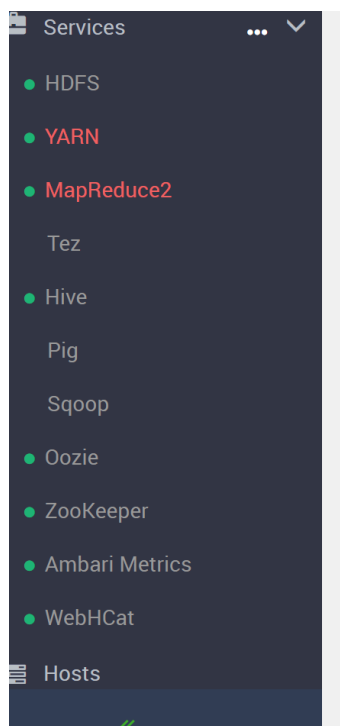
Browse

```
1 select Attrition, Department, JobSatisfaction,MonthlyIncome from employee_sales;
```

default Tables(3)

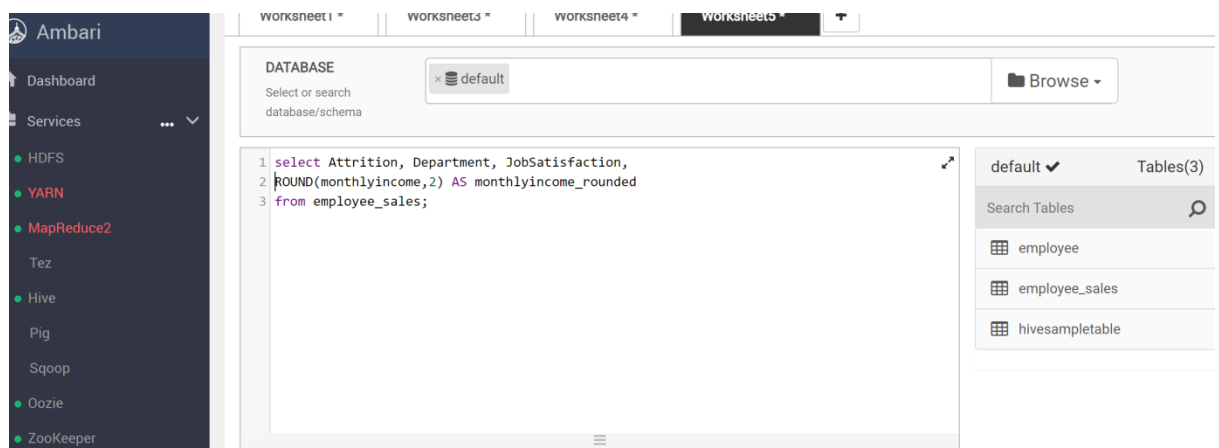
Search Tables

- employee
- employee_sales
- hivesampletable



Filter col				
attrition	department	jobsatisfaction	monthlyincome	
Yes	Sales	4	5993	
No	Research & Development	2	5130	
Yes	Research & Development	3	2090	
No	Research & Development	3	2909	
No	Research & Development	2	3468	
No	Research & Development	4	3068	
No	Research & Development	1	2670	
No	Research & Development	3	2693	
No	Research & Development	3	9526	
No	Research & Development	3	5237	
No	Research & Development	2	2426	

1- Rounding the data found in the “MonthlyIncome” column to the nearest \$1000.



Services					Filter col	✕
attrition	department	jobsatisfaction	monthlyincome			
Yes	Sales	4	5993			
No	Research & Development	2	5130			
Yes	Research & Development	3	2090			
No	Research & Development	3	2909			
No	Research & Development	2	3468			
No	Research & Development	4	3068			
No	Research & Development	1	2670			
No	Research & Development	3	2693			
No	Research & Development	3	9526			
No	Research & Development	3	5237			
No	Research & Development	2	2426			

attrition	department	jobsatisfaction	monthlyincome
Yes	Sales	4	5993
No	Research & Development	2	5130
Yes	Research & Development	3	2090
No	Research & Development	3	2909
No	Research & Development	2	3468
No	Research & Development	4	3068
No	Research & Development	1	2670
No	Research & Development	3	2693
No	Research & Development	3	9526
No	Research & Development	3	5237
No	Research & Development	2	2426

3) Filtering the data to only look at those items in the “Sales Department”.

YARN

MapReduce2

Tez

Hive

Pig

Sqoop

Oozie

ZooKeeper

Ambari Metrics

WebHCat

Hosts

Worksheet1 * ✕

Worksheet3 * ✕

Worksheet4 * ✕

Worksheet5 * ✕

+

DATABASE

✕ default

Browse

1 insert overwrite table employee_sales

2

3 select Attrition, Department, JobSatisfaction,monthlyincome

4 from employee_sales

5 where Department like "Sales";

6

default ✓ Tables(3)

Search Tables

employee

employee_sales

hivesampletable

DATABASE

Select or search database/schema

default

Browse

1 insert overwrite table employee_sales

2

3 select Attrition, Department, JobSatisfaction,monthlyincome

4 from employee_sales

5 where Department like "Sales";

6

default

Tables(3)

Search Tables

employee

employee_sales

hivesampletable

Dashboard

Services

HDFS

YARN

MapReduce2

Tez

Hive

Pig

Sqoop

Oozie

ZooKeeper

Ambari Metrics

WebHCat

Filter col

✕

attrition	department	jobsatisfaction	monthlyincome
Yes	Sales	4	5993
No	Sales	4	15427
Yes	Sales	1	3407
No	Sales	2	6825
No	Sales	1	18947
Yes	Sales	4	2086
Yes	Sales	3	2683
No	Sales	4	2014
No	Sales	1	5376
No	Sales	3	8726

4-Ordering data by “JobSatisfaction” from highest to lowest

Ambari

Dashboard

Services

HDFS

YARN

MapReduce2

Tez

Hive

Pig

Sqoop

Oozie

Select or search database/schema

1

2 insert overwrite table employee_sales

3 select Attrition, Department, JobSatisfaction,monthlyincome

4 from employee_sales

5 order by JobSatisfaction DESC;

6

default ✓ Tables(3)

Search Tables

employee

employee_sales


hivesampletable

✓ Execute

Save As

Insert UDF

Visual Explain

 Ambari

Dashboard

Services

● HDFS

● YARN

● MapReduce2

Tez

● Hive

Pig

Sqoop

● Oozie

● ZooKeeper

● Ambari Metrics

attrition	department	jobsatisfaction	monthlyincome
No	Human Resources	4	6347
No	Sales	4	4648
No	Research & Development	4	3452
No	Sales	4	5337
Yes	Research & Development	4	2340
No	Research & Development	4	7491
No	Research & Development	4	2406
No	Sales	4	4950
No	Research & Development	4	2781
No	Sales	4	5220
No	Sales	4	5772
No	Sales	4	2400