

5. Write an SQL query that perform outer join the two tables (the tables from 2.a and 2.b)

The screenshot shows the MySQL Workbench interface. The SQL editor contains the following query:

```
1 select e1.StoreCode as StoreCode,
2 e1.Date as Date,
3 e1.TotalSales as TotalSales,
4 e1.TotalAmount as TotalAmount,
5 j1.city as city,
6 j1.size as size,
7 j1.latitude as latitude,
8 j1.longitude as longitude,
9 j1.population as population,
10 j1.state as state
11 from excitable1 as e1
12 left join
13 jsontable1 as j1
14 on e1.StoreCode = j1.code
15
16 union all
17
18 select e1.StoreCode as StoreCode,
```

The Result Grid shows the output of the query, displaying columns: StoreCode, Date, TotalSales, TotalAmount, city, size, latitude, longitude, population, and state. The data includes rows for StoreCodes 'zd4QT', 'oZIEy', 'HydF', 'lmGR', and 'bzdhj'.

The Output pane shows the execution details: 1 21:59:51 select e1.StoreCode as StoreCode, e1.Date as Date, e1.TotalSales as TotalSales, e1.TotalAmount as TotalAm... 1990998 row(s) returned. Duration / Fetch: 0.015 sec / 7.469 sec.

The screenshot shows the MySQL Workbench interface. The SQL editor contains the following query:

```
23 union all
24
25 select e1.StoreCode as StoreCode,
26 e1.Date as Date,
27 e1.TotalSales as TotalSales,
28 e1.TotalAmount as TotalAmount,
29 j1.city as city,
30 j1.size as size,
31 j1.latitude as latitude,
32 j1.longitude as longitude,
33 j1.population as population,
34 j1.state as state
35 from excitable1 as e1
36 right join
37 jsontable1 as j1
38 on e1.StoreCode = j1.code;
39
40
```

The Result Grid shows the output of the query, displaying columns: StoreCode, Date, TotalSales, TotalAmount, city, size, latitude, longitude, population, and state. The data includes rows for StoreCodes 'zd4QT', 'oZIEy', 'HydF', 'lmGR', and 'bzdhj'.

The Output pane shows the execution details: 2 22:21:23 select distinct count(*) from jsontable1 1 row(s) returned 0.016 sec / 0.000 sec; 3 22:21:39 select distinct count(*) from excitable1 1 row(s) returned 0.203 sec / 0.000 sec; 4 22:25:13 select e1.StoreCode as StoreCode, e1.Date as Date, e1.TotalSales as TotalSales, e1.TotalAmount as TotalAm... 1990998 row(s) returned 0.015 sec / 7.469 sec.

a) How many rows are produced by the SQL query?

Answer: 1998998

The screenshot shows the MySQL Workbench interface. The SQL editor contains the following query:

```

30 j1.size as size,
31 j1.latitude as latitude,
32 j1.longitude as longitude,
33 j1.population as population,
34 j1.state as state
35 from exceltable1 as e1
36 right join
37 jsontable1 as j1
38 on e1.StoreCode = j1.code;
39
40
41
42 SELECT count(*) FROM jsontable1 LEFT OUTER JOIN "exceltable1" ON jsontable1.code = "exceltable1".StoreCode
43 union all
44 SELECT count(*) FROM jsontable1 RIGHT OUTER JOIN "exceltable1" ON jsontable1.code = "exceltable1".StoreCode
45 ;
46
47

```

The Results tab shows the following output:

count(*)
998998
1000000

The Action Output tab shows the following messages:

#	Time	Action	Message	Duration / Fetch
3	22:21:39	select distinct count(*) from exceltable1	1 row(s) returned	0.203 sec / 0.000 sec
4	22:25:13	select e1.StoreCode as StoreCode, e1.Date as Date, e1.TotalSales as TotalSales, e1.TotalAmount as TotalAm...	1998998 row(s) returned	0.015 sec / 7.469 sec
5	22:29:05	SELECT count(*) FROM jsontable1 LEFT OUTER JOIN "exceltable1" ON jsontable1.code = "exceltable1".Store...	2 row(s) returned	2.031 sec / 0.000 sec

b) How many rows are there that did not have a lookup row?

Answer: 1002

The screenshot shows the Apache Spark SQL IDE interface. The pipeline consists of three steps:

```

graph LR
    CSVTableInput[CSVTable input] --> DatabaseLookup[Database lookup]
    DatabaseLookup --> FinalTableOutput[FinalTable output]

```

The Execution Results tab shows the following data:

#	Stepname	Copypn	Read	Written	Input	Output	Updated	Rejected	Errors	Active	Time	Speed (r/s)	Input/output
1	CSVTable input	0	0	1000000	1000000	0	0	0	0	Finished	8mn 58s	1,257	-
2	Database lookup	0	1000000	1000000	998998	0	0	0	0	Finished	9mn 3s	1,239	-
3	FinalTable output	0	1000000	1000000	0	1000000	0	0	0	Finished	9mn 4s	1,238	-