

Assignment – Day 17

-Sarthak Niranjana Kulkarni (Maverick)

- sarthakkul2311@gmail.com

- (+91) 93256 02791

28/11/2024 (Thursday)

Practice on Delta Tables: -

1. Loading and Displaying Data from Delta Table in Azure Databricks

→ `spark.table("default.export")`

`data = spark.read.format("delta").load("dbfs:/user/hive/warehouse/export")`

`data.show()`

```
▶ data: pyspark.sql.dataframe.DataFrame = [id: long, firstName: string ... 6 more fields]
```

3	Quyen	Marlen	Dome	F	1970-10-11 04:00:00	957-57-8246	53417
4	Coralie	Antonina	Marshal	F	1990-04-11 04:00:00	963-39-4885	94727
5	Terrie	Wava	Bonar	F	1980-01-16 05:00:00	964-49-8051	79908
6	Chassidy	Concepcion	Bourthouloume	F	1990-11-24 05:00:00	954-59-9172	64652
7	Geri	Tambra	Mosby	F	1970-12-19 05:00:00	968-16-4020	38195
8	Patria	Nancy	Arstall	F	1985-01-02 05:00:00	984-76-3770	102053
9	Terese	Alfredia	Tocque	F	1967-11-17 05:00:00	967-48-7309	91294
10	Wava	Lyndsey	Jeandon	F	1963-12-30 05:00:00	997-82-2946	56521
11	Sophie	Emerita	Hearn	F	1979-09-17 04:00:00	977-66-4483	90920
12	Jodie	Tabetha	Laneham	F	1959-01-31 05:00:00	923-24-9769	90634
13	Marietta	Mandi	Yansons	F	1974-02-19 04:00:00	900-34-8083	93162
14	Caridad	Maire	Snelle	F	1960-09-26 04:00:00	992-11-7062	38859
15	Yasmine	Meg	Edworthy	F	1960-01-29 05:00:00	922-12-9862	76220
16	Chan	Jani	Hartas	F	1986-12-05 05:00:00	995-51-3115	75050
17	Evangeline	Wanetta	Casserley	F	1961-09-29 04:00:00	926-61-3526	62814
18	Elnora	Kecia	Lipman	F	1980-02-14 05:00:00	950-23-9739	71350
19	Adelle	Kathryn	Grigoriev	F	1978-11-14 05:00:00	923-23-5984	60600
20	Mica	Zandra	Challens	F	1973-11-24 05:00:00	918-66-1232	51071

only showing top 20 rows

2. Writing DataFrame to Delta Tables in Azure Databricks

→ from pyspark.sql import SparkSession

```
spark = SparkSession.builder.appName('Delta Table Write').getOrCreate()
```

```
data = spark.createDataFrame([
```

```
    (1, "Alice", 1000),
```

```
    (2, "Bob", 2000),
```

```
    (3, "Charlie", 3000)
```

```
], ["id", "name", "salary"])
```

Write the DataFrame as Delta tables

```
data.write.format('delta').saveAsTable("mydata_delta", mode="overwrite")
```

```
data.write.format('delta').saveAsTable("mydata")
```

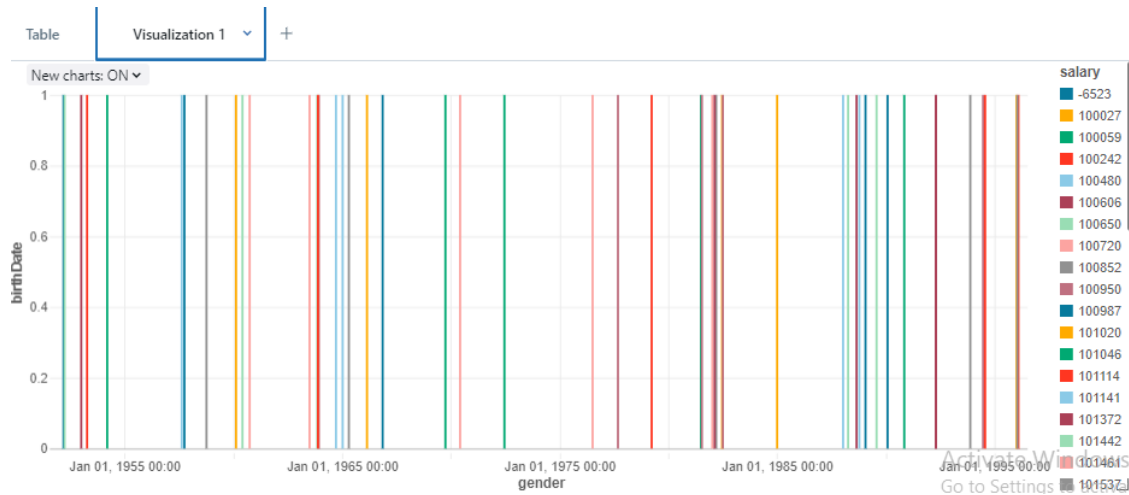
3. Loading and Displaying Data from Delta Table in Databricks

→ spark.table("default.export")

```
data = spark.read.format("delta").load("dbfs:/user/hive/warehouse/export")
```

```
data.display()
```

Table	Visualization 1								
id	first Name	middle Name	last Name	gender	birthDate	ssn	salary		
48	Carie	Serrina	Waryk	F	1957-02-11T05:00:00.000+00:00	980-30-1636	52549		
49	Pansy	Suzie	Shrieves	F	1991-05-24T04:00:00.000+00:00	910-16-4444	73811		
50	Malissa	Amada	McRill	F	1958-05-02T04:00:00.000+00:00	954-65-2862	115845		
51	Blythe	Carri	Crinkley	F	1973-03-08T05:00:00.000+00:00	998-29-3568	112662		
52	Marketta	Rachele	Borg-Bartolo	F	1975-10-10T04:00:00.000+00:00	981-60-8178	66712		
53	Curtis	Temple	Sponton	F	1980-07-23T04:00:00.000+00:00	997-70-4544	74767		
54	Mariceia	Eiane	Heinel	F	1972-12-21T05:00:00.000+00:00	931-49-8310	67801		
55	Delpha	Michelina	Biggadike	F	1957-09-30T04:00:00.000+00:00	987-71-9556	104262		
56	Bernita	Kathaleen	McIan	F	1971-02-20T05:00:00.000+00:00	982-68-4659	48781		
57	Jesusa	Cherise	Parysowna	F	1988-03-14T05:00:00.000+00:00	992-77-3435	111979		
58	Chung	Dian	Dautry	F	1998-01-12T05:00:00.000+00:00	946-10-3058	47190		
59	Vickie	Ranae	Saddleton	F	1967-04-01T05:00:00.000+00:00	955-57-2783	52410		
60	Sena	Particia	Stittie	F	1956-01-10T05:00:00.000+00:00	971-10-8053	72745		



Summary on Delta Tables: -

In Azure Databricks, Delta tables are used to store data in a structured format that supports efficient querying and data management. The first block of code shows how to read data from an existing Delta table stored in the Databricks metastore or from a path in the DBFS (Databricks File System). Using the `spark.read.format("delta")` method, we load the data from the Delta table into a `DataFrame` and display it. This process allows us to view the content of the Delta table, which is stored in a structured format for analysis.

Next, we see how to create and write a new Delta table from a `DataFrame`. The `data.write.format("delta").saveAsTable("mydata")` command writes the data into a new Delta table named "mydata". This code demonstrates the ability to save a `DataFrame` into a Delta table, making it accessible for future queries and operations. We can specify the mode (like overwrite) to control how existing data is handled when writing the new data.

Finally, the `data.display()` method is used to show the contents of the newly written Delta table in a Databricks notebook. Overall, Delta tables provide a powerful and efficient way to store and manage data in Databricks, with built-in support for ACID transactions, versioning, and schema enforcement. This makes them ideal for data analysis and machine learning tasks where data integrity and fast querying are crucial.