All PySpark Date-Related Scenario-Based Interview Questions

Karthik Kondpak

Swipe for more

1. Extracting Year:

Given a column event_date (format: yyyy-MM-dd), extract only the year for each row.

Sample data:

event_date: ["2023-04-15", "2022-11-30", "2021-08-25"]

2.Date Difference Calculation:

Calculate the difference (in days) between two date columns start_date and end_date.

Sample data:

start_date: ["2023-01-01", "2023-03-15"]
end_date: ["2023-02-01", "2023-03-20"]

3. Filter Records Based on Date:

Filter records where event_date is after 2023-06-01.

Sample data:

event_date: ["2023-05-15", "2023-07-20", "2023-06-05"]

4. Add Days to Date:

Add 30 days to each date in the order_date column.

Sample data:

order_date: ["2023-01-10", "2023-06-15", "2023-07-30"]

5.Find the Maximum Date:

Determine the latest date from a column payment_date.

Sample data:

payment_date: ["2023-02-15", "2023-06-25", "2023-01-10"]

6.Truncate Date to First Day of Month:

Truncate the sale_date to the first day of its respective month.

Sample data:

sale_date: ["2023-04-12", "2023-07-23", "2023-08-05"]

7. Group by Year:

Group records by year extracted from the column transaction_date.

Sample data:

transaction_date: ["2023-06-12", "2022-11-09", "2021-04-01"]

8.Filter Records Within a Date Range:

Filter records where visit_date is between 2023-01-01 and 2023-05-01.

Sample data:

visit_date: ["2023-02-15", "2023-06-01", "2023-03-20"]

9. Day of Week Extraction:

Extract the day of the week from attendance_date.

Sample data:

attendance_date: ["2023-08-11", "2023-07-25", "2023-09-01"]

10. Check Leap Year:

Identify if each date in birth_date falls in a leap year.

Sample data:

birth_date: ["2020-03-01", "2019-12-15", "2024-02-29"]

11.Convert String to Date:

Convert a string column arrival_time (format: dd-MM-yyyy) to date format.

Sample data:

arrival_time: ["15-04-2023", "20-08-2023", "01-12-2023"]

12. Calculate Week Number:

For each date in shipment_date, calculate the week number of the year.

Sample data:

shipment_date: ["2023-02-15", "2023-08-01", "2023-12-25"]

Swipe for more

13. Find Records from the Last 7 Days:

Identify all records where log_date is within the last 7 days from the current date.

Sample data:

log_date: ["2023-08-08", "2023-08-11", "2023-08-15"]

14. Format Date as String:

Format the booking_date as dd/MM/yyyy.

Sample data:

booking_date: ["2023-07-12", "2023-09-15", "2023-05-30"]

15. Find the First and Last Record by Date:

Find the first and last record based on the created_at date.

Sample data:

created_at: ["2023-02-10", "2023-07-01", "2023-03-15"]

16. Difference Between Dates in Months:

Calculate the difference between two dates in months (start_month and end_month).

Sample data:

start_month: ["2023-01-01", "2023-04-01"]
end_month: ["2023-06-01", "2023-07-01"]

17. Convert UTC to Local Time:

Convert the utc_timestamp column from UTC to a local timezone (e.g., IST).

Sample data:

utc_timestamp: ["2023-08-14T12:00:00Z", "2023-08-14T18:30:00Z"]

18. Find Holidays:

Check if dates in the holiday_date column are public holidays (e.g., 2023-01-01, 2023-12-25).

Sample data:

holiday_date: ["2023-01-01", "2023-11-24", "2023-12-25"]

19. Round Time to Nearest Hour:

Round the meeting_time column to the nearest hour.

Sample data:

meeting_time: ["2023-08-14T12:35:00", "2023-08-14T18:25:00"]

---

20. Extract Quarter:

Extract the quarter of the year from the invoice_date.

Sample data:

invoice_date: ["2023-03-15", "2023-06-01", "2023-10-20"]

If you find this helpful, like and share it with your friends

https://www.seekhobigdata.com/