parquet (orc) - DateTimeExceedLimit

Last updated by | Veena Pachauri | Mar 8, 2023 at 11:22 PM PST

Issue:

When reading data out from parquet (or ORC) file, if datetime is involved, Julian Calendar is used for dates before Oct 15, 1582. Compared with Gregorian Calendar, which is used at data factory runtime, there are gaps. You can find more information here:
https://en.wikipedia.org/wiki/Proleptic_Gregorian_calendar#Difference_between_Julian_and_proleptic_Gregorian_calendar_dates ⬀

for example, if **0001-01-01 00:00:00Z** is stored in parquet file, it is a Julian date, in Gregorian Calendar, it is BC 0001-12-30 00:00:00 (or to be more comprehensive 0000-12-30 00:00:00), and this exceeds the date limit. The error message will be:

Failure happened on 'Sink' side. ErrorCode=ParquetDateTimeExceedLimit,'Type=Microsoft.DataTransfer.Common.Shared.HybridDeliveryException,Message=The Ticks value '-621357696000000000' for the datetime column must be between valid datetime ticks range -621355968000000000 and 2534022144000000000.,Source=Microsoft.DataTransfer.Richfile.ParquetTransferPlugin

Resolution:

Change the date column value large than Oct 15, 1582 in source.