# Create a dim date dimension for the gold layer

Add the following script to your notebook to create a date dimension

|  |
| --- |
| from pyspark.sql.functions import col, explode, sequence, to\_date, year, month, dayofmonth  # Define the begin and end dates as parameters  begin\_date = "20220101"  end\_date = "20221231"  database\_name = "my\_database"  table\_name = "date\_dimension"  # Create a DataFrame with a sequence of dates between the begin and end dates  df\_dates = (spark.range(0, (to\_date(col(end\_date)) - to\_date(col(begin\_date))).cast("int"))  .withColumn("date", to\_date(col(begin\_date)) + col("id"))  .select(col("date").alias("full\_date")))  # Extract the year, month, and day from the full date and convert to integers  df\_dates = (df\_dates  .withColumn("year", year(col("full\_date")).cast("int"))  .withColumn("month", month(col("full\_date")).cast("int"))  .withColumn("day", dayofmonth(col("full\_date")).cast("int")))  # Add a key column in the format of yyyymmdd  df\_dates = (df\_dates  .withColumn("key", col("full\_date").cast("int")))  # Write the DataFrame to a Delta table  (df\_dates.write  .format("delta")  .option("overwriteSchema", "true")  .mode("overwrite")  .saveAsTable(f"{database\_name}.{table\_name}"))  # Show the resulting DataFrame  df\_dates.show() |