

ERROR YANG DIALAMI PADA SAAT MEMBUAT PROJECT

Deskripsi Error

Error ini menunjukkan bahwa program mencoba mengakses sebuah file atau direktori yang tidak ditemukan pada lokasi (path) yang ditentukan.

```
File "/opt/airflow/scripts/bronze_ingest.py", line 18, in run_bronze_ingestion
    with open(path, "w") as f:
        ~~~~~~
FileNotFoundError: [Errno 2] No such file or directory: '/opt/airflow/data/bronze/flights_20260101083020.json'
[2026-01-01, 08:30:20 UTC] {taskinstance.py:1206} INFO - Marking task as FAILED. dag_id=flights_ops_medallion_pipe,
[2026-01-01, 08:30:20 UTC] {standard_task_runner.py:110} ERROR - Failed to execute job 4 for task bronze_ingest ([Errno 2] No such file or directory: '/opt/airflow/data/bronze/flights_20260101083020.json')
[2026-01-01, 08:30:20 UTC] {local_task_job_runner.py:243} INFO - Task exited with return code 1
[2026-01-01, 08:30:20 UTC] {taskinstance.py:3503} INFO - 0 downstream tasks scheduled from follow-on schedule check
[2026-01-01, 08:30:20 UTC] {local_task_job_runner.py:222} ▲▲▲ Log group end
```

Direktori tujuan belum dibuat. File yang diharapkan belum ada saat proses dijalankan. Path file yang digunakan tidak sesuai dengan struktur direktori di environment (misalnya container Docker).

Deskripsi Error

Error ini terjadi karena method mkdir() dari modul pathlib.Path menerima argumen yang tidak dikenali.

```
File "/opt/airflow/scripts/silver_transform.py", line 17, in run_silver_transform
    silver_path.mkdir(parent=True, exist_ok=True)
TypeError: Path.mkdir() got an unexpected keyword argument 'parent'
[2026-01-01, 10:07:45 UTC] {taskinstance.py:1206} INFO - Marking task as FAILED. dag_id=flights_ops_medallion_pipe, task_id=silver_transform, run_id=manual_2026-01-01
[2026-01-01, 10:07:45 UTC] {standard_task_runner.py:110} ERROR - Failed to execute job 10 for task silver_transform (Path.mkdir() got an unexpected keyword argument 'parent')
[2026-01-01, 10:07:45 UTC] {local_task_job_runner.py:243} INFO - Task exited with return code 1
[2026-01-01, 10:07:45 UTC] {taskinstance.py:3503} INFO - 0 downstream tasks scheduled from follow-on schedule check
[2026-01-01, 10:07:45 UTC] {local_task_job_runner.py:222} ▲▲▲ Log group end
```

Kesalahan ini disebabkan oleh penulisan nama parameter yang tidak sesuai dengan API resmi pathlib. Pada pathlib.Path.mkdir(), parameter yang valid adalah:

- parents (bentuk jamak)
- exist_ok

Deskripsi Error

Error ini menunjukkan adanya **ketidaksesuaian jumlah elemen** antara struktur data yang sudah ada dengan data baru yang ingin diterapkan.

```
File "/home/airflow/.local/lib/python3.12/site-packages/pandas/core/internals/base.py", line 85, in _validate_set_axis
    raise ValueError(
ValueError: Length mismatch: Expected axis has 17 elements, new values have 18 elements
[2026-01-01, 10:15:13 UTC] {taskinstance.py:1206} INFO - Marking task as FAILED. dag_id=flights_ops_medallion_pipe, task_
[2026-01-01, 10:15:13 UTC] {standard_task_runner.py:110} ERROR - Failed to execute job 12 for task silver_transform (Leng
[2026-01-01, 10:15:13 UTC] {local_task_job_runner.py:243} INFO - Task exited with return code 1
[2026-01-01, 10:15:13 UTC] {taskinstance.py:3503} INFO - 0 downstream tasks scheduled from follow-on schedule check
[2026-01-01, 10:15:13 UTC] {local_task_job_runner.py:222} ▲▲ Log group end
```

Objek data (biasanya DataFrame Pandas) memiliki 17 elemen pada axis tertentu (umumnya kolom). Namun, kode saya mencoba menetapkan 18 nilai baru pada axis tersebut. Pandas menolak operasi ini karena jumlah elemen harus sama persis. Disini saya sesuaikan df columns nya menjadi 17.

Deskripsi Error

Error ini menunjukkan bahwa kode mencoba mengakses atau memproses kolom yang tidak ada di dalam sebuah DataFrame Pandas. Tepatnya pada colomns ‘on_gound’.

```
File "/home/airflow/.local/lib/python3.12/site-packages/pandas/core/apply.py", line 639, in normalize_dictlike_arg
    raise KeyError(f"Column(s) {list(cols)} do not exist")
KeyError: "Column(s) ['on_ground'] do not exist"
[2026-01-01, 11:13:53 UTC] {taskinstance.py:1206} INFO - Marking task as FAILED. dag_id=flights_ops_medallion_pipe, ta
[2026-01-01, 11:13:53 UTC] {standard_task_runner.py:110} ERROR - Failed to execute job 21 for task gold_aggregate ("Co
[2026-01-01, 11:13:53 UTC] {local_task_job_runner.py:243} INFO - Task exited with return code 1
[2026-01-01, 11:13:53 UTC] {taskinstance.py:3503} INFO - 0 downstream tasks scheduled from follow-on schedule check
[2026-01-01, 11:13:53 UTC] {local_task_job_runner.py:222} ▲▲ Log group end
```

Sebelumnya saya menggunakan “geo_altitude” pada df = df_raw di file silver_transform.py(proses sebelum gold). Disini saya Ganti dan sesuaikan dataframe file di silver_transform.py menjadi ‘on_ground’ agar sesuai dengan tugas pada layer gold_aggregate.

The screenshot shows the Snowflake SQL editor interface. On the left, there's a sidebar with 'My Workspace' and '+ Add new'. The main area has tabs for 'Home' and 'Untitled.sql'. The code editor contains the following SQL:

```

1 CREATE DATABASE FLIGHTS;
2
3 CREATE SCHEMA KPI;
4
5 CREATE TABLE FLIGHT_KPIs (
6     window_start TIMESTAMP,
7     origin_country TEXT,
8     total_flights INT,
9     avg_velocity FLOAT,
10    on_ground INT,
11    load_time TIMESTAMP DEFAULT CURRENT_TIMESTAMP,
12    PRIMARY KEY (window_start, origin_country)

```

Below the code editor is a 'Results' panel with the message: 'Cannot perform CREATE TABLE. This session does not have a current database. Call 'USE DATABASE' or use a qualified name.' There are 0 rows returned in 22ms.

Session Snowflake kamu belum memilih database aktif, sehingga Snowflake tidak tahu tabel itu harus dibuat di database mana. Disini saya belum menentukan warehouse, database dan schemas yang ingin saya gunakan.

Python mencoba memanggil fungsi atau nama strftime yang tidak dikenali, karena strftime bukan fungsi global, strftime adalah method milik objek datetime. Namun di kode tersebut, strftime dipanggil tanpa objek datetime, Python lalu menganggap strftime sebagai nama variabel atau fungsi, dan karena tidak pernah didefinisikan, muncullah NameError.

```

File "/opt/airflow/scripts/load_gold_to_snowflake.py", line 14, in load_gold_to_snowflake
    execution_date = context["data_interval_start"],strftime("%Y-%m-%d-%H:%M:%S")
                           ^^^^^^
NameError: name 'strftime' is not defined
[2026-01-01, 14:49:34 UTC] {taskinstance.py:1206} INFO - Marking task as FAILED. dag_id=flights_ops_medallion_pipe,
[2026-01-01, 14:49:34 UTC] {standard_task_runner.py:110} ERROR - Failed to execute job 49 for task load_gold_to_snow
[2026-01-01, 14:49:34 UTC] {local_task_job_runner.py:243} INFO - Task exited with return code 1
[2026-01-01, 14:49:34 UTC] {taskinstance.py:3503} INFO - 0 downstream tasks scheduled from follow-on schedule check
[2026-01-01, 14:49:34 UTC] {local_task_job_runner.py:222} ▲▲▲ Log group end

```

Disini saya salah mengetik seharusnya menggunakan titik(.) sebelumnya saya menggunakan koma(,)

Menyatakan bahwa timestamp tidak dikenali karena tidak ada hubungan antara tanggal dengan waktu.

```
errors.ProgrammingError: 100035 (22007): 01c17140-0001-3386-0000-000089994791: Timestamp '2026-01-01-14:00:00' is not recognized
[  UTC] {taskinstance.py:1206} INFO - Marking task as FAILED. dag_id=flights_ops_medallion_pipe, task_id=load_gold_to_snowflake, r
[  UTC] {standard_task_runner.py:110} ERROR - Failed to execute job 53 for task load_gold_to_snowflake (100035 (22007): 01c17140-0
[  UTC] {local_task_job_runner.py:243} INFO - Task exited with return code 1
[  UTC] {taskinstance.py:3503} INFO - 0 downstream tasks scheduled from follow-on schedule check
[  UTC] {local_task_job_runner.py:222} ▲▲ Log group end
```

Pada kode `execution_date = context["data_interval_start"].strftime("%Y-%m-%d-%H:%M:%S")` di layer load to csv seharusnya tidak dikasih tanda hubung hanya spasi saja