

Dag in prefect

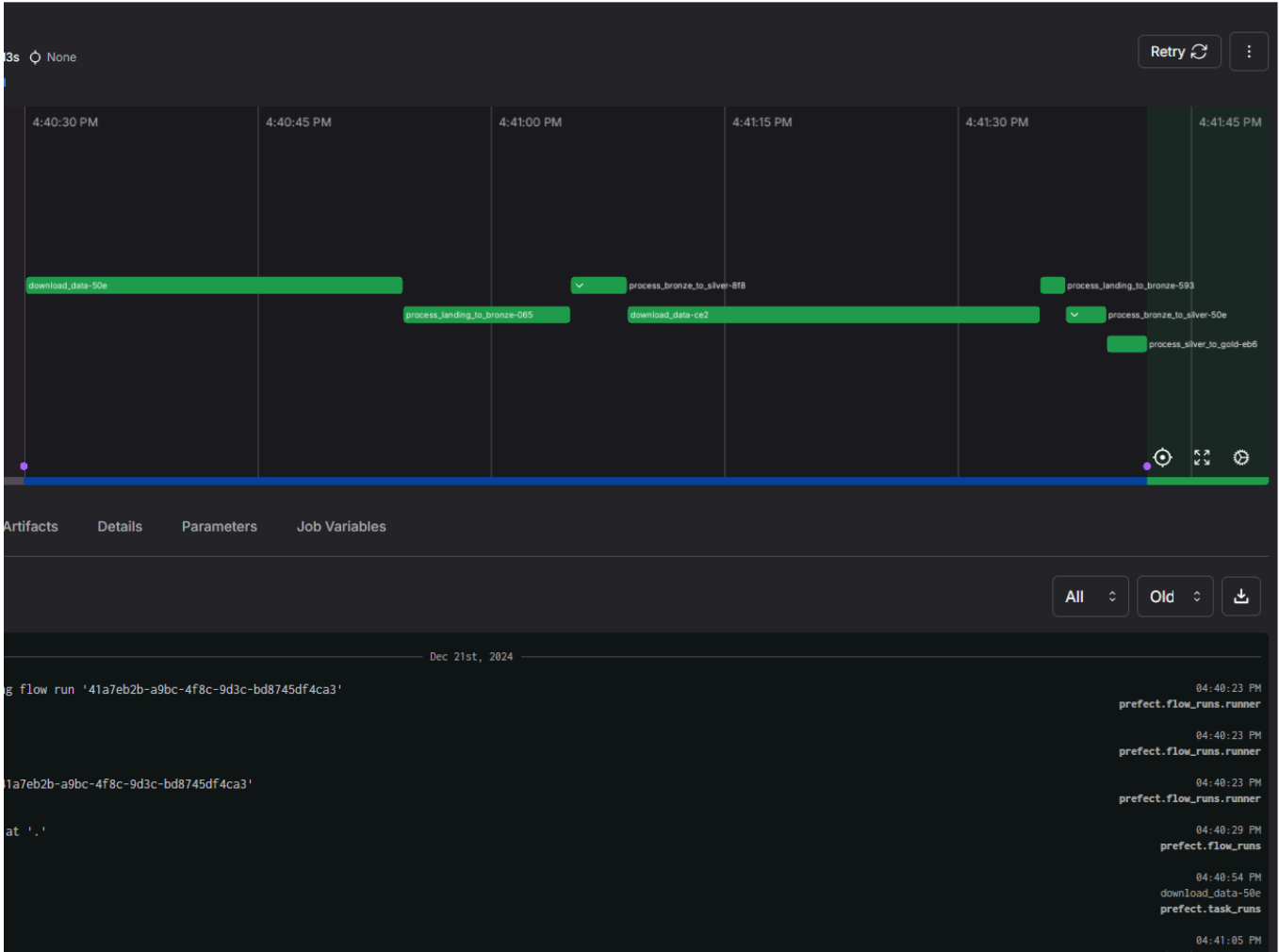


Table bronze

```
Welcome  fp_part2_wsl.py X
fp_part2_wsl.py > ...
30 def process_landing_to_bronze(table):
31     spark = get_spark_session("LandingtoBronze")
32
33     local_path = f"{table}.csv"
34     output_path = f"/tmp/bronze/{table}"
35
36     # Read the CSV into a Spark DataFrame
37     df = spark.read.csv(local_path, header=True, inferSchema=True)
38
39     # Ensure output directory exists
40     os.makedirs(output_path, exist_ok=True)
41
42     # Write the DataFrame to Parquet
43     df.write.mode("overwrite").parquet(output_path)
44     print(f>Data saved to {output_path}")
45
46     # Optional: Re-read and verify
47     df = spark.read.parquet(output_path)
48     df.show()#truncate=False)
49
50
51     spark.stop()
52
```

PROBLEMS OUTPUT DEBUG CONSOLE **TERMINAL** PORTS 2

Data saved to /tmp/bronze/athlete_bio

athlete_id	name	sex	born	height	weight	country	country_noc	description	special_notes
59950	Bruno Jutzeler	Male	1 January 1946	182	95	Switzerland	SUI	Bruno Jutzeler ha...	Brother of Peter ...
75279	Danie Jacobs	Male	10 April 1904	NULL	NULL	South Africa	RSA	Danie Jacobs was ...	NULL
126505	Zhuldyz Eshimova	Female	2 January 1988	152	48	Kazakhstan	KAZ	NULL	Listed in Olympia...
39978	Jiří Havlis	Male	16 November 1932	NULL	NULL	Czechoslovakia	TCH	NULL	NULL
131108	Houleye Ba	Female	17 July 1992	170	55	Mauritania	MTN	NULL	NULL
108875	Esther Termens	Female	2 November 1984	170	65	Spain	ESP	NULL	NULL
109352	Angeliki Karapatakis	Female	19 February 1975	170	58	Greece	GRE	NULL	Wife of Nikolaos ...
60520	Anatoly Roshchin	Male	10 March 1932	187	100	Soviet Union	URS	Anatoly Roshchin, ...	NULL
21994	Joseph Perotiaux	Male	8 January 1883	NULL	NULL	France	FRA	NULL	NULL
39338	Szczepan Grajczyk	Male	11 December 1931	176	81	Poland	POL	NULL	NULL
7821	José Antonio Colón	Male	17 August 1948	173	71	Puerto Rico	PUR	NULL	NULL
33550	Bruno Carabetta	Male	27 July 1966	172	72	France	FRA	NULL	NULL
9090	Heinz Rodinger	Male	21 February 1941	182	75	Austria	AUT	NULL	NULL
115688	Hanna Westrin	Female	25 November 1991	168	64	Sweden	SWE	NULL	Listed in Olympia...
32075	Erich Herrmann	Male	21 May 1914	NULL	NULL	Germany	GER	Field handball wa...	NULL
78625	John Paul Jones	Male	15 October 1890	179	66	United States	USA	John Paul Jones w...	NULL
65376	Renelda Swan	Male	12 December 1957	178	70	Bermuda	BER	Personal Best: 40...	NULL
69646	Hildrun Laufer-Claus	Female	13 May 1939	172	66	Germany	GER	In 1962, Hildrun ...	Wife of Peter Laufer
1339	Beibis Mendoza	Male	20 June 1974	165	48	Colombia	COL	NULL	Listed in Olympic...
126608	Cammie Adams	Female	11 September 1991	173	65	United States	USA	NULL	Listed in Olympia...

only showing top 20 rows

Table silver

```
fp_part2_wsl.py > ...
30 def process_landing_to_bronze(table):
31     spark = get_spark_session("LandingtoBronze")
32
33     local_path = f"{table}.csv"
34     output_path = f"/tmp/bronze/{table}"
35
36     # Read the CSV into a Spark DataFrame
37     df = spark.read.csv(local_path, header=True, inferSchema=True)
38
39     # Ensure output directory exists
40     os.makedirs(output_path, exist_ok=True)
41
42     # Write the DataFrame to Parquet
43     df.write.mode("overwrite").parquet(output_path)
44     print(f>Data saved to {output_path}")
45
46     # Optional: Re-read and verify
47     df = spark.read.parquet(output_path)
48     df.show()#truncate=False)
49
50     spark.stop()
51
52
53
54 @task
55 def process_landing_to_bronze(table):
```

PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS 2

Data saved to /tmp/silver/athlete_bio

athlete_id	name	sex	born	height	weight	country	country_noc	description	special_notes
68467	yann quentrec	male	17 february 1962	190	75	france	fra	personal best: 40...	listed in olympia...
76395	sven strömberg	male	22 january 1911	176	73	sweden	swe	sven strömberg ha...	listed in olympia...
31357	scott keswick	male	3 march 1970	162	57	united states	usa	scott keswick com...	listed in olympia...
74358	robin tait	male	14 april 1940	191	131	new zealand	nzl	personal best: dt...	listed in olympia...
80922	doris de agostini	female	28 april 1958	178	59	switzerland	sui	doris de agostini...	listed in olympia...
83813	jindřich balcar	male	22 march 1950	181	85	czechoslovakia	tch	jindřich balcar w...	brother of jarosl...
79387	novica čanović	male	29 november 1961	193	71	yugoslavia	yug	personal best: hj...	listed in olympia...
68166	henri dartigues	male	17 september 1902	168	57	france	fra	henri dartigues w...	listed in olympia...
75695	sugath thilakaratne	male	30 july 1973	170	60	sri lanka	sri	personal best: 40...	listed in olympia...
47757	sandra völker	female	1 april 1974	180	75	germany	ger	sandra völker was...	listed in olympia...
64589	robyn woodhouse	female	26 july 1943	174	60	australia	aus	personal best: hj...	listed in olympia...
86361	michael hadschieff	male	5 october 1963	185	75	austria	aut	arguably the best...	listed in olympia...
70153	eckart berkes	male	9 february 1949	180	70	west germany	frg	eckart berkes sta...	listed in olympia...
115044	marvin anderson	male	12 may 1982	175	69	jamaica	jam	personal bests: 2...	listed in olympia...
64580	norma thrower	female	5 february 1936	166	57	australia	aus	norma thrower had...	listed in olympia...
88378	tomasz ścigaczewski	male	18 november 1978	188	85	poland	pol	personal best: 11...	listed in olympia...
69461	frank sando	male	14 march 1931	171	52	great britain	gbr	frank sando ran t...	listed in olympia...
14626	zoltán halász	male	23 july 1960	176	68	hungary	hun	between 1970 and ...	brother of lászló...
114639	kerron clement	male	31 october 1985	188	86	united states	usa	personal bests: 4...	listed in olympia...
29573	aleksej demjanov	male	9 december 1973	162	59	croatia	cro	aleksej demjanov ...	listed in olympia...

Table gold

```
16:41:42.251 | INFO | prefect - Stopping temporary server on http://127.0.0.1:8197
+-----+-----+-----+-----+-----+-----+-----+
| sport | medal | sex | country_noc | avg_height | avg_weight | timestamp |
+-----+-----+-----+-----+-----+-----+-----+
| rowing | gold | male | nzl | 189.80645161290323 | 88.80645161290323 | 2024-12-21 16:41:... |
| athletics | bronze | male | gre | 188.66666666666666 | 82.33333333333333 | 2024-12-21 16:41:... |
| athletics | gold | female | can | 170.5 | 58.166666666666664 | 2024-12-21 16:41:... |
| athletics | NULL | female | ned | 171.93037974683546 | 61.18354430379747 | 2024-12-21 16:41:... |
| canoe marathon | NULL | male | gbr | 183.0 | 79.0 | 2024-12-21 16:41:... |
| cycling track | NULL | male | fin | 179.47619047619048 | 73.0 | 2024-12-21 16:41:... |
| canoe sprint | NULL | female | mex | 176.0 | 61.666666666666664 | 2024-12-21 16:41:... |
| tennis | NULL | female | rou | 174.375 | 62.583333333333336 | 2024-12-21 16:41:... |
| table tennis | NULL | female | tun | 163.42857142857142 | 57.857142857142854 | 2024-12-21 16:41:... |
| football | NULL | male | sui | 181.12 | 77.16 | 2024-12-21 16:41:... |
| beach volleyball | NULL | female | ger | 180.86363636363637 | 70.22727272727273 | 2024-12-21 16:41:... |
| equestrian eventing | NULL | female | mex | 160.0 | 52.0 | 2024-12-21 16:41:... |
| cycling track | NULL | male | mas | 169.24 | 66.04 | 2024-12-21 16:41:... |
| archery | NULL | female | jpn | 164.36170212765958 | 58.57446808510638 | 2024-12-21 16:41:... |
| diving | NULL | female | jpn | 157.24242424242425 | 51.24242424242424 | 2024-12-21 16:41:... |
| cycling bmx racing | NULL | female | can | 160.0 | 60.0 | 2024-12-21 16:41:... |
| tennis | NULL | male | can | 188.92857142857142 | 82.53571428571429 | 2024-12-21 16:41:... |
| equestrian eventing | NULL | male | pol | 175.53030303030303 | 69.59090909090909 | 2024-12-21 16:41:... |
| sailing | NULL | male | ukr | 182.96428571428572 | 77.75 | 2024-12-21 16:41:... |
| tennis | NULL | male | per | 175.5 | 72.0 | 2024-12-21 16:41:... |
+-----+-----+-----+-----+-----+-----+-----+
only showing top 20 rows
```