

打包文件主要包含：hl-TPCDI (hanlin的文档)、tools (生成TPC-DI数据的工具)、sf_3 (里面有数据, 23个sql task脚本和1个python task脚本以及定义他们的dag.py, 其他sf代码相似不放了)、create_image.py (airflow统计数据画图的), 还有本pdf。

上届GitHub: <https://github.com/risg99/tpc-di-benchmark>, 可参考报告和readme

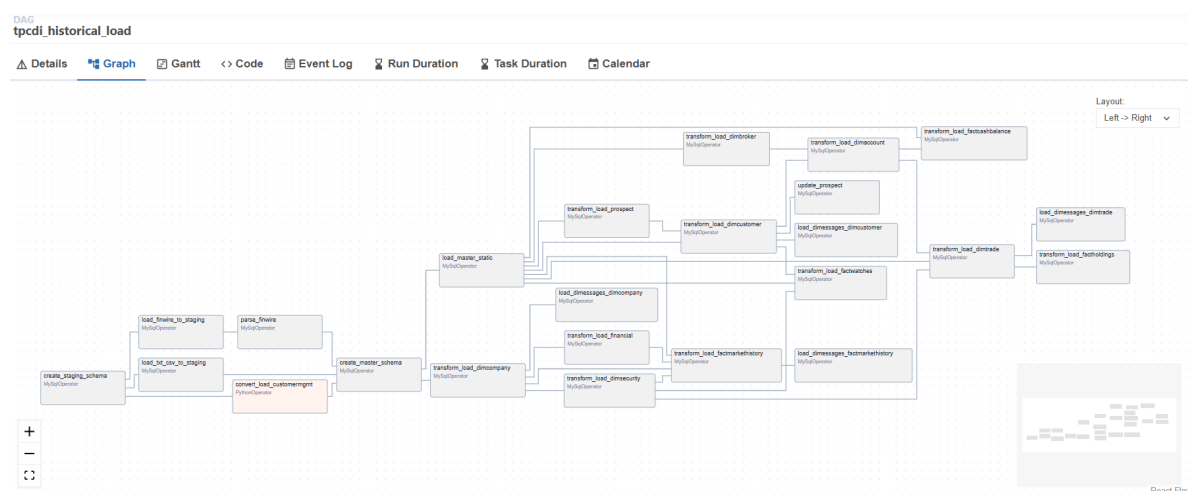
各工具版本: MySQL 8.0.29(就是mariadb)、Docker Desktop 4.34.2

这里提一嘴, hanlin和这个组最大不同就是hanlin区分了静态表和非静态, 前者是数据从头到尾不改变的, 所以她用一个sql全打包生成了, 所以我们的DAG dependency略微不同, 具体详见hanlin文档, dependency详见dag.py

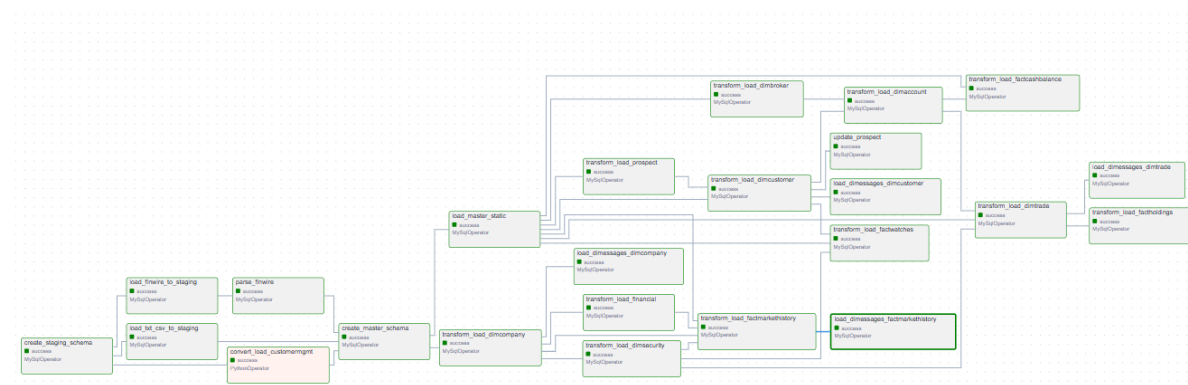
sf_9数据生成, 替换hanlin文档sf24

```
All work done
Data generation finished successfully
AuditTotalRecordsSummaryWriter - TotalRecords for Batch1: 14338913
AuditTotalRecordsSummaryWriter - TotalRecords for Batch2: 60698
AuditTotalRecordsSummaryWriter - TotalRecords for Batch3: 60685
AuditTotalRecordsSummaryWriter - TotalRecords all Batches: 14460296 1057193.74 records/second
Statistics
=====
Overall time      0h:00m:13s:677ms
Generated        858.8 MiB
Speed            62.8 MiB/s
```

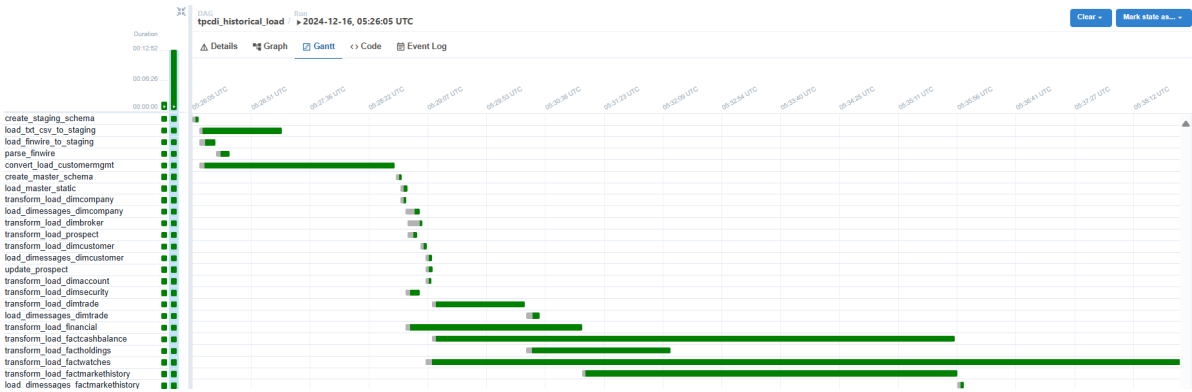
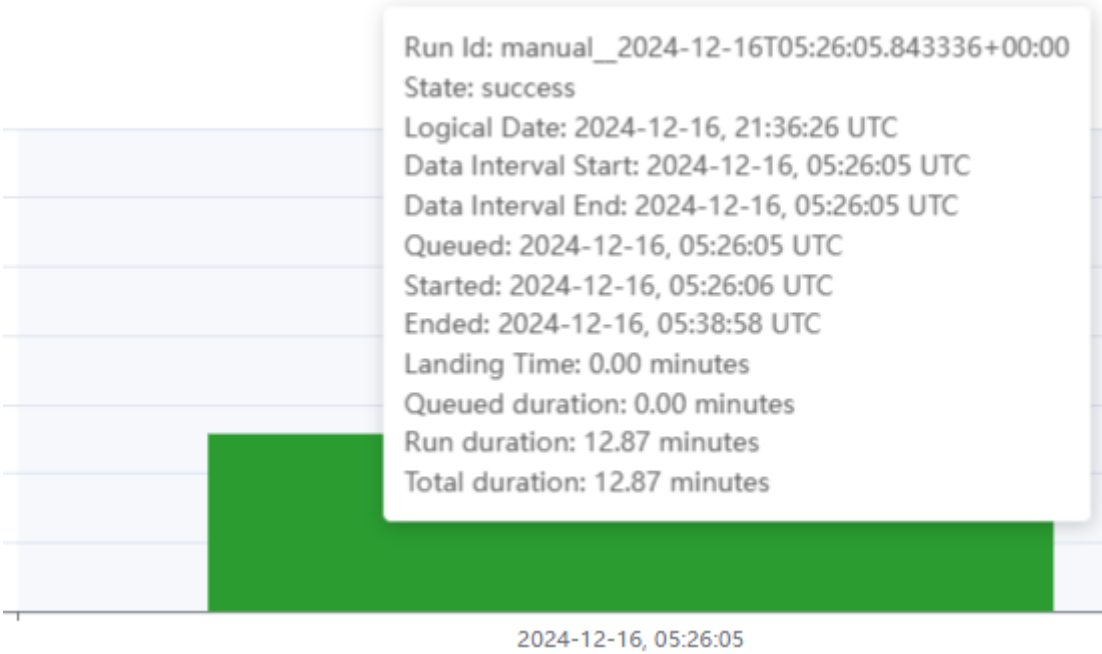
DAG graph



完成示意图



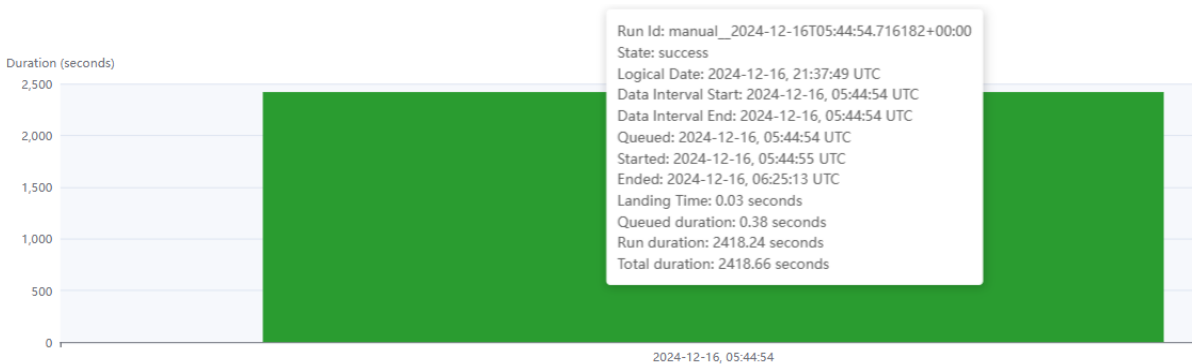
sf_3 run duration & gantt & task duration graphs

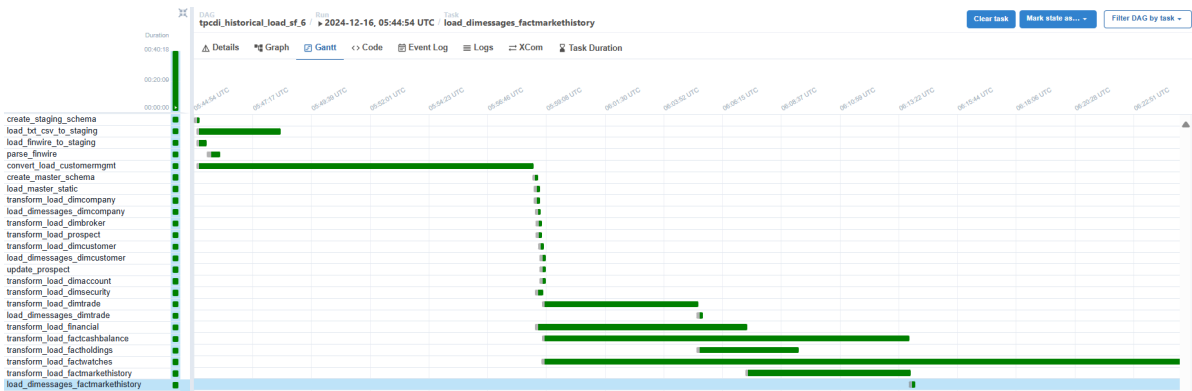
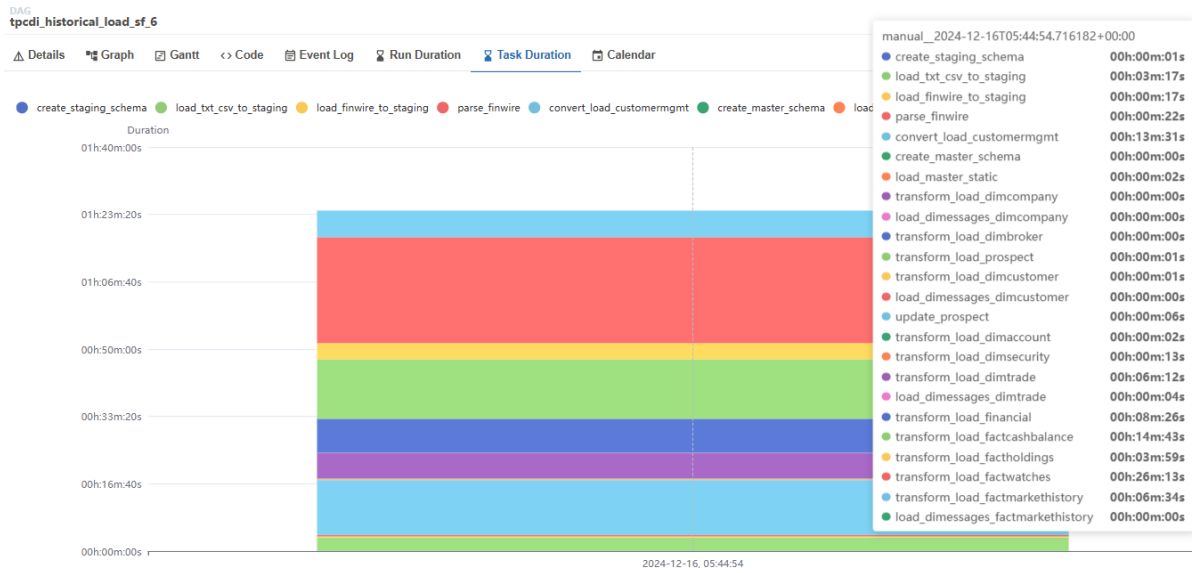


跑了两遍

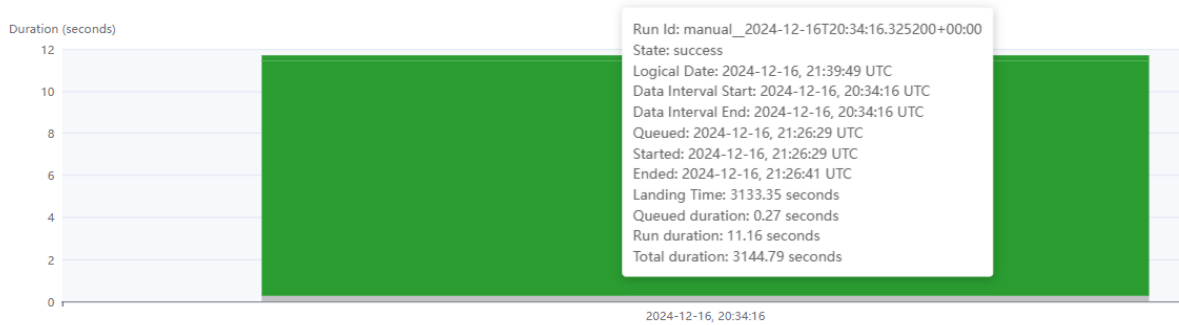


sf_6



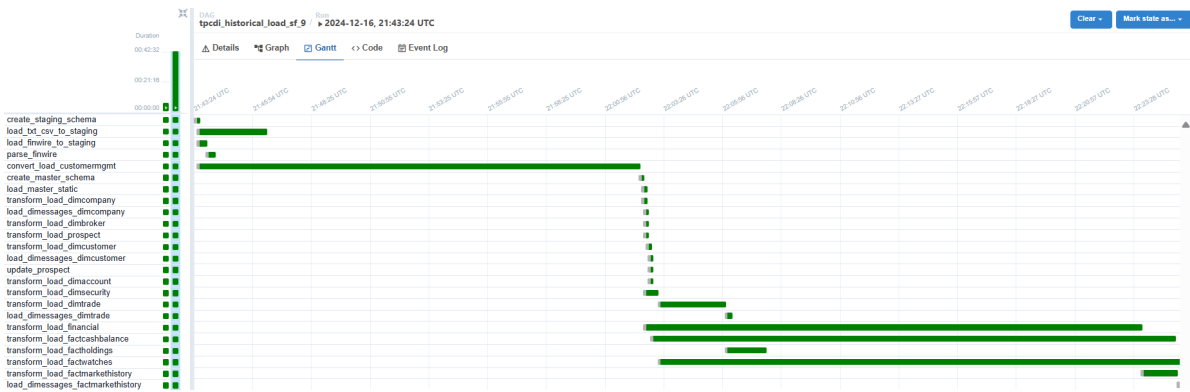
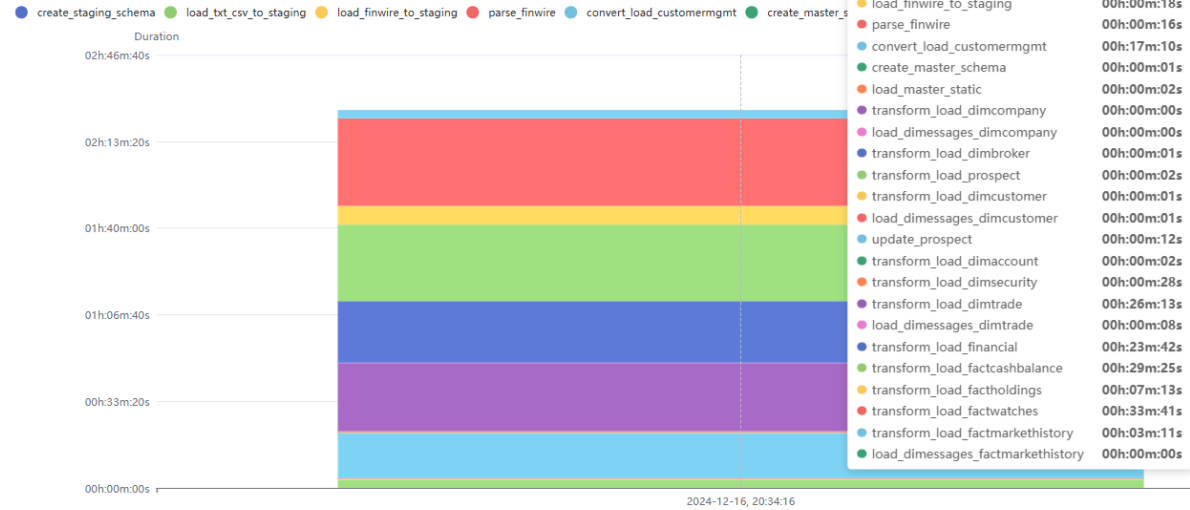


sf_9



DAG

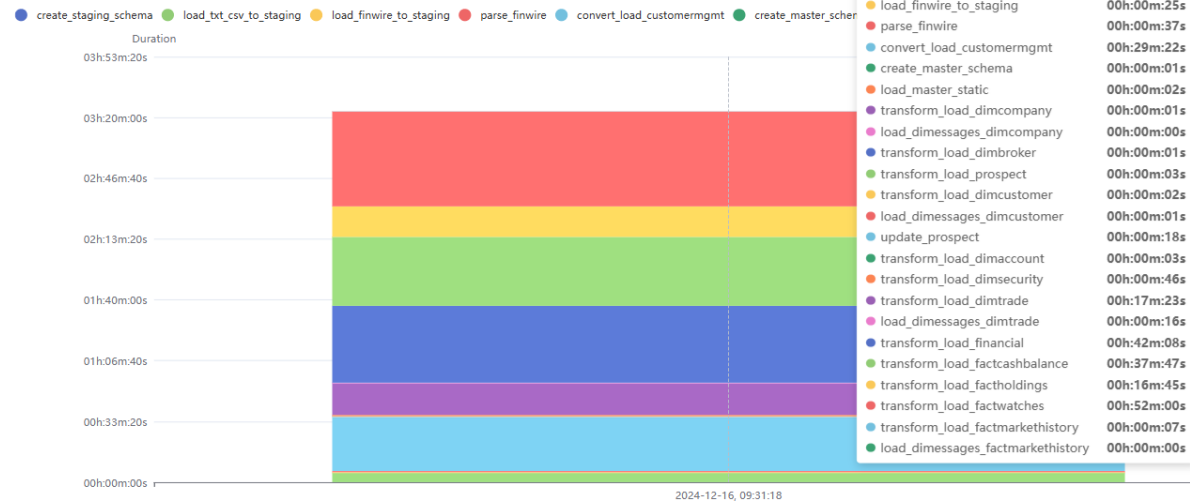
tpcdi_historical_load_sf_9

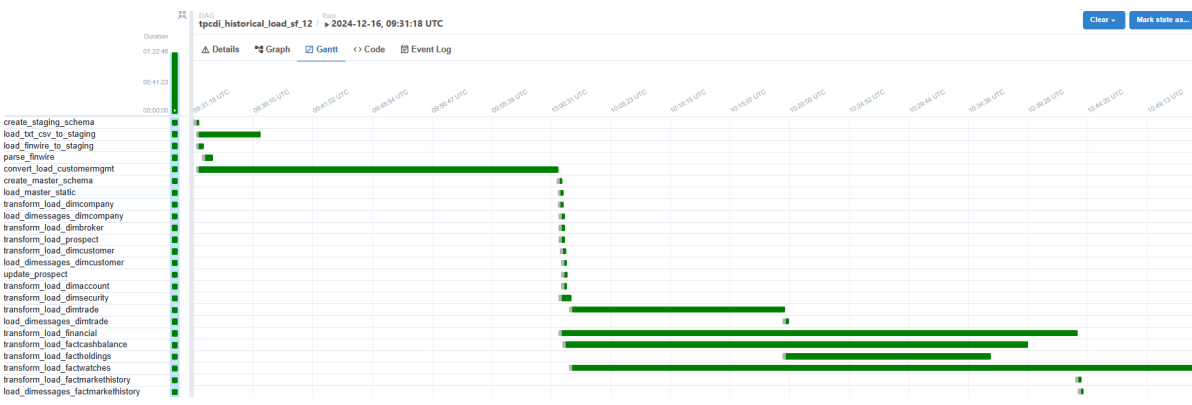
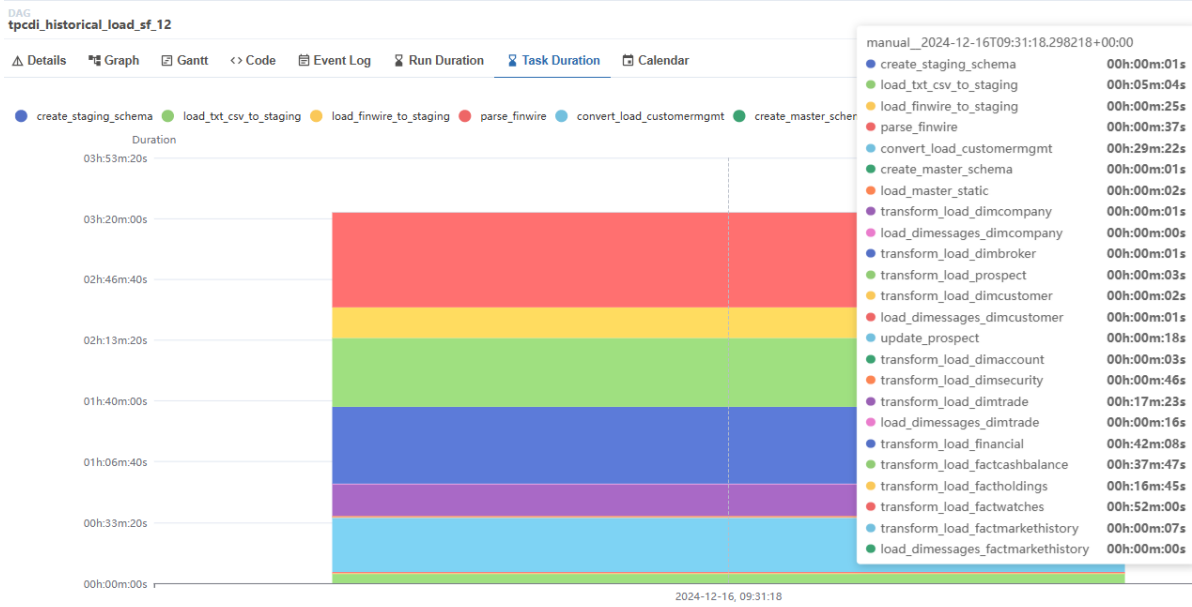
[Details](#) [Graph](#) [Gantt](#) [Code](#) [Event Log](#) [Run Duration](#) [Task Duration](#) [Calendar](#)

sf_12

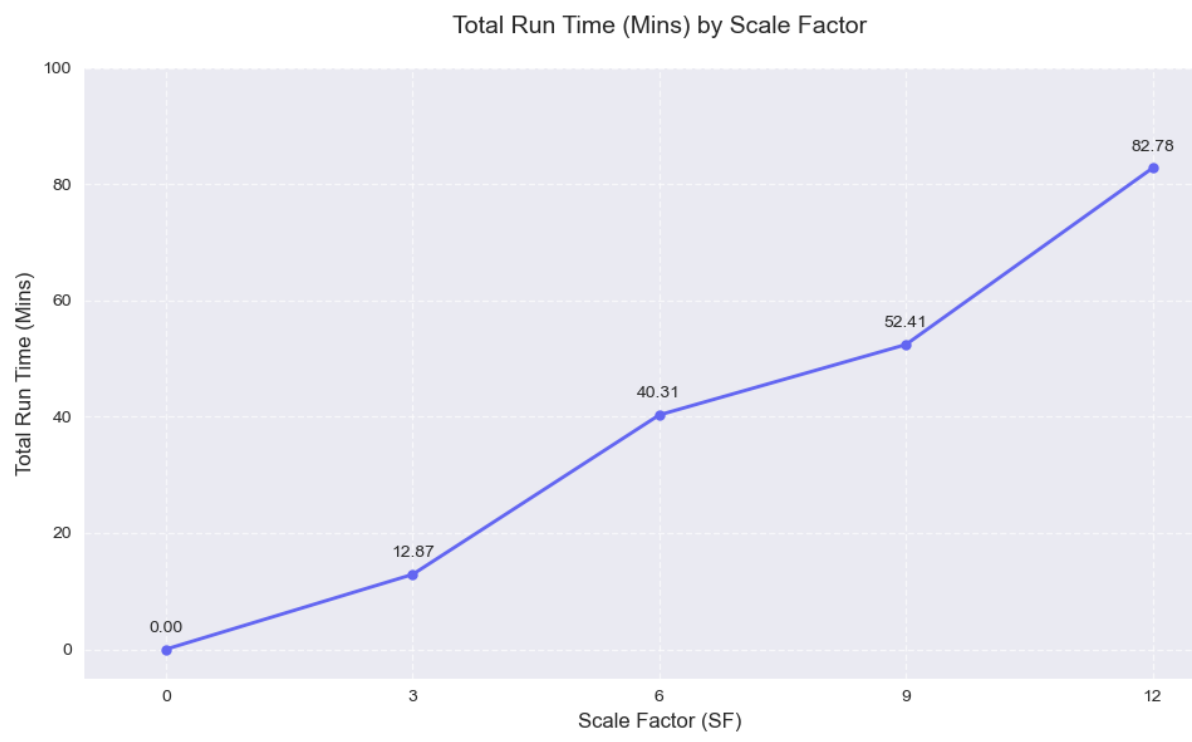
DAG

tpcdi_historical_load_sf_12

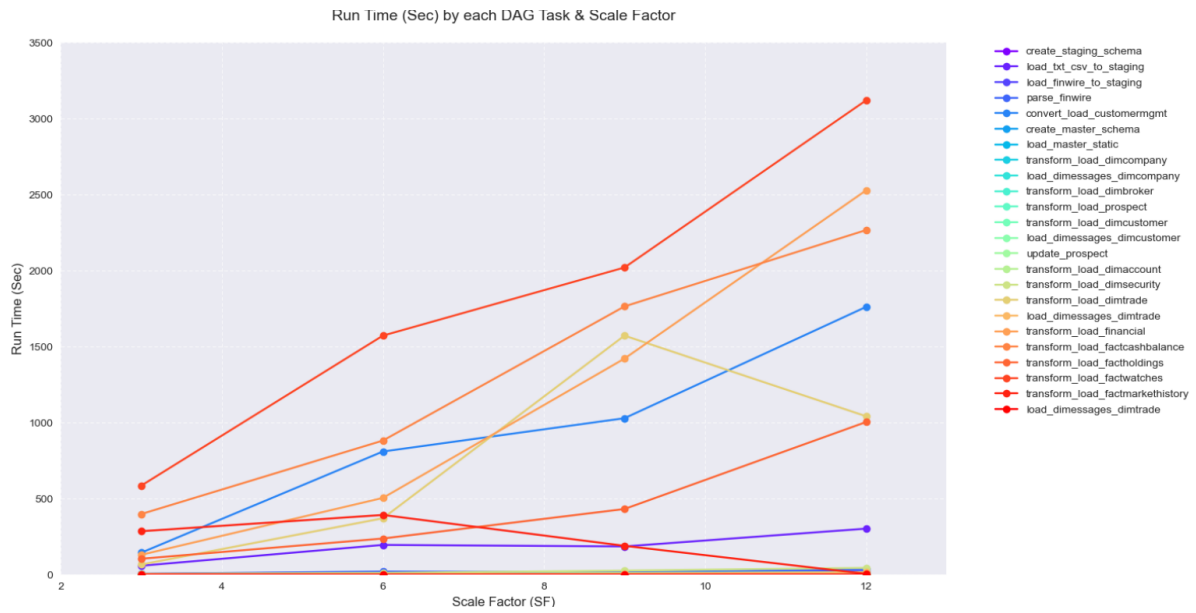
[Details](#) [Graph](#) [Gantt](#) [Code](#) [Event Log](#) [Run Duration](#) [Task Duration](#) [Calendar](#)



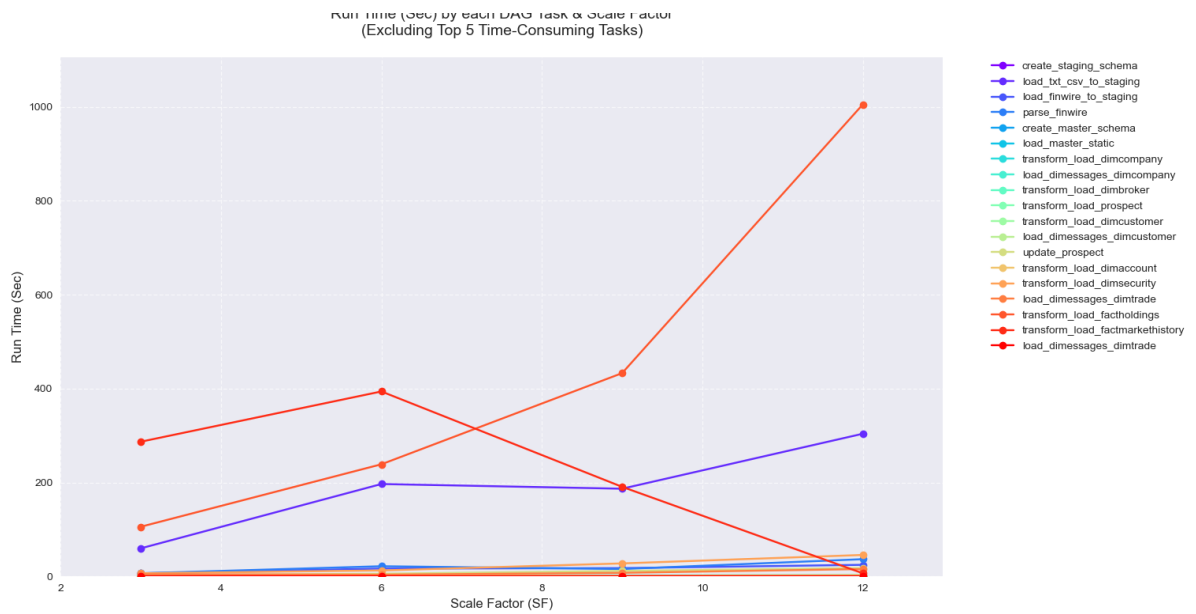
total_run_time



run_time_by_task



run_time_by_task_exclude_top_5



我贴心的为你画了表格，里面的数据详见create_image.py

TASK ID	SF 3	SF 6	SF 9	SF 12
create_staging_schema	1	1	1	1
load_txt_csv_to_staging	60	197	187	304
load_finwire_to_staging	7	17	18	25
parse_finwire	7	22	16	37
convert_load_customermgmt	146	811	1030	1762
create_master_schema	0	0	1	1
load_master_static	2	2	2	2
transform_load_dimcompany	0	0	0	1
load_dimessages_dimcompany	3	0	0	0

TASK ID	SF 3	SF 6	SF 9	SF 12
transform_load_dimbroker	1	0	1	1
transform_load_prospect	2	1	2	3
transform_load_dimcustomer	1	1	1	2
load_dimessages_dimcustomer	1	0	1	1
update_prospect	3	6	12	18
transform_load_dimaccount	1	2	2	3
transform_load_dimsecurity	7	13	28	46
transform_load_dimtrade	68	372	1573	1043
load_dimessages_dimtrade	5	4	8	16
transform_load_financial	132	506	1422	2528
transform_load_factcashbalance	400	883	1765	2267
transform_load_factholdings	106	239	433	1005
transform_load_factwatches	587	1573	2021	3120
transform_load_factmarkethistory	287	394	191	7
load_dimessages_dimtrade	0	0	0	0