

## Short Summary

Last week, we integrated and explored multiple datasets using Python's pandas library, identifying key entities, attributes, and relationships. Data inconsistencies such as mismatched types and null values were noted.

Initially, we defined the business process as “**every time a user places an order,**” but this week we corrected it to “**every product for every order a user makes.**” The focus shifted from `order_id` to a newly defined `line_id`. The conceptual model was revised accordingly, and Kimball's dimensional modeling was still applied to implement the logical and physical models in SQL.

- We designed the ETL/ELT pipeline in Kestra, creating a Git repository with `seeds/` and `models/` directories similar to a **tutorial we found of Kestra UI**.
- Source files are converted to CSV with python
- Using `io.kestra.plugin.dbt.cli.DbtCLI` Kestra can find and read the files on our repository.
- CSV files are staged using queries on file names **`stg_*.sql`** where we rename fields, cast data types, and standardize format.
- Results are transformed according to the schema using queries on file names **`*_dim.sql`** and **`*_fact.sql`**.
- Finally, we export the dim and fact tables as *outputFiles on Kestra UI* using python.

Before we move on, we need consultation to finalize project decisions, including confirming assumptions, where to store output, task modularization, removal of hardcoding, and whether datasets will be updated.

## Evidence of progress

shopzada

Jump to...

Ctrl/Cmd + K

Execute

Overview

Topology

Executions

Edit

Revisions

Triggers

Logs

Metrics

Dependencies

Concurrency

Audit Logs

Flow Code

No-code

Topology

Documentation

Files

Blueprints

Playground

Actions

Saved

Flow Code

Topology

```

4 tasks:
5   - id: dbt
6     tasks:
7       - id: dbt_build
8         profiles: |
9           dbt:
10             type: io.kestra.plugin.scripts.python.Script
11             outputFiles:
12               - "*/.csv"
13             taskRunner:
14               type: io.kestra.plugin.scripts.runner.docker.Docker
15               containerImage: ghcr.io/kestra-io/duckdb:latest
16             script: |
17               import duckdb
18               import pandas as pd
19
20               conn = duckdb.connect(database='dbt.duckdb', read_only=False)
21
22               tables_query = "SELECT table_name FROM information_schema.
23                               tables WHERE table_schema = 'main';"
24
25               tables = conn.execute(tables_query).fetchall()
26
27               # Export each table to CSV, only if the table name ends with
28               # 'dim' or 'fact'
29               for table_name in tables:
30                 table_name = table_name[0]
31                 if table_name.endswith('dim') or table_name.endswith
32                 ('fact'):
33                   query = f"SELECT * FROM {table_name}"
34                   df = conn.execute(query).fetchdf()
35                   df.to_csv(f"{table_name}.csv", index=False)
36
37               conn.close()

```

AI Copilot

```

graph TD
    dbt --> clone_repository
    clone_repository --> prepare_for_staging
    prepare_for_staging --> dbt_build
    dbt_build --> export

```

main

1 Branch

0 Tags

Add file

Code

FRADMGit Create staff_dim.sql		e9bf9b3 · yesterday	62 Commits
models	Create staff_dim.sql	yesterday	
seeds	Delete seeds/schema.yml	yesterday	
dbt_project.yml	Create dbt_project.yml	3 days ago	
profiles.yml	Create profiles.yml	3 days ago	

Flows

shopZada.shopZada / Executions

6Zmgu6tBvFY3979cHUVUdn

kestra

Flows

Apps

Executions

Logs

Tests

Namespaces

KV Store

Secrets

Blueprints

Plugins

Administration

Kestra

Overview

Tasks

dbt

0 items

clone\_repository

1 items

prepare\_for\_staging

4 items

dbt\_build

4 items

vars

exitCode

taskRunner

outputFiles

Preview

Displayed content has been truncated

Row count

100

Encoding

UTF-8

View

Auto view

Enforce editor view

```

staff_id,staff_name,staff_job_level,staff_street,staff_state,staff_city,staff_company,staff_c
STAFF000065,Randall Bergstrom,Intermediate,376 Land chester,Texas,Omahla,CO 64 95 3488
STAFF000396,Christian Hessel,Intermediate,945 West Camp shire,New Mexico,SAN JOY,can
STAFF004923,Eduardo Fadel,entry,89 Expressway tom, Rhode Island,Corpus Christi,Alabama,320
STAFF001819,Jordi Gleicher,entry,728 Centers burgh,Virginia,Bakersfield,Timor-Leste,649258
STAFF003616,Price Hintz,intermediate,1720 North Skyway burgh,Alabama,Scottsdale,PA,Palestine
STAFF002363,Garfield Legros,intermediate,596 Way tom,Virginia,San Bernardino,Antarctica,349
STAFF002193,Shane Volsen,entry,2124 North Plains borough,New Hampshire,Tomb, Thailand,36136
STAFF005152,Dudley Goodwin,intermediate,54835 Main way,Towa,Colorado Springs,Lesotho,2615
STAFF002359,Curt Thiel,entry,8114 Tale furt,New York,Anahle,Gambia,288756959,2921-11-24 1
STAFF004929,Kiley Monahan,intermediate,2222 Land bury,South Dakota,Indragana,Morgolia,115
STAFF002975,Coleman Haag,intermediate,24698 Row side,North Carolina,San Francisco,North Mac
STAFF000683,Jaquean Metcalf,entry,3253 South Ridge mound,Minnesota,Kansas,South Georgia and
STAFF005582,Evelly Connelly,intermediate,942 North Drives tom,Indiana,Chula Vista,Morocco,588
STAFF006567,Dustin Turcotte,intermediate,4224 East Kent,New Hampshire,Aurora,Poland,7188
STAFF006172,Tevin Wellmann,entry,138 North Covess mound,Vest Virginia,Scottsdale,Slovakia,718
STAFF002869,Grady Howell,entry,743 Turquoise side,Maine,Arlington,Eritrea,685229888,2021-01-
STAFF004136,Zachary Krick,intermediate,41581 Port mill borough,Nebraska,Vienna,Poland,806
STAFF000365,Rick Murfin,entry,2990 Branch burgh,California,Kansas,Viet Nam,355041427,2023-0
STAFF004195,Bert Cole,intermediate,3807 Lake Viaduct way,Virginia,Long Beach,Portugal,2256
STAFF005663,Sedrick Welter,senior,6293 New Mall borough,Nevada,Tolado,American Samoa,1853223
STAFF004840,Willa Dick's,entry,9932 East Ben chomter,Minnesota,Scottsdale,Guibouti,19724265
STAFF0062223,Citlalli Munoz,intermediate,97573 Port Creek tom,Wisconsin,Norfolk,Cabo Ve
STAFF005583,Vernon Mann,senior,36123 New Court burgh,illinois,Denver,Morilo,968492646
STAFF004693,Joell Wellmann,entry,989 North Passage port,Indiana,Aurora,Macao,207844472,7821
STAFF001493,James Johnston,senior,3319 North Lake bury,Missouri,Tolado,Nigeria,145597379

```