



Project: Adventure Works

Foundation First !!!

Four Key Questions

- I. Where do we consolidate our data ? > [Storage](#)
- II. How will we get it there ? > [Ingestion](#)
- III. How will we clean it up? > [Transformation](#)
- IV. How will we analyze it? > [Reporting](#)



BRIAN GWAYI





Data Stack

Popular Options

Storage > [Snowflake](#), [BigQuery](#), [s3](#), Redshift

Ingestion > [Airbyte](#), [Airflow](#), Fivetran

Transformation > [dbt](#)

Reporting > [Tableau](#), Power BI, [Looker](#), Superset

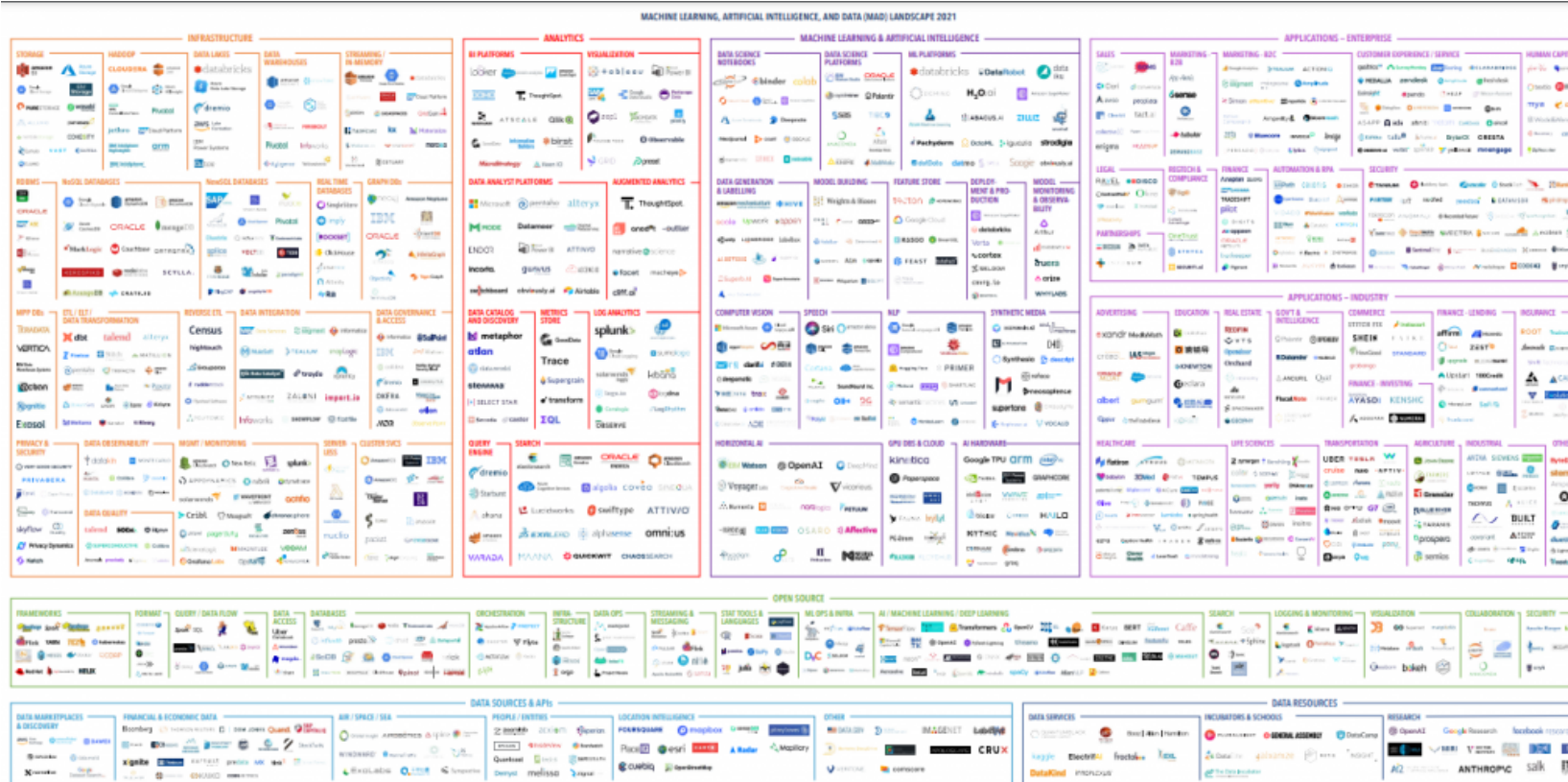
N/B This is not an exhaustive list of options.

BRIAN GWAYI



Modern Data Stack Ecosystem 2024

The right tools for building robust data stack architecture will be based on Combination of budget, skillset, data sources and preferences.



Architecture Design

Source



Ingestion

Airflow



Python



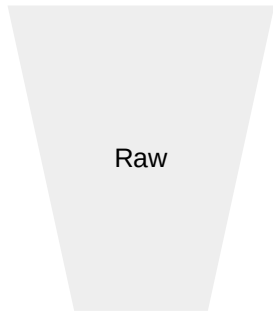
Storage
BigQuery



Transformation
dbt



dbt



Model

Model

Model

Reporting
Looker



Project : E-commerce

01

Storage/Database

Setting Google [BigQuery](#)

02

Ingestion

Setting up [Apache Airflow](#)

Writing elt Python script

Orchestrate data pipeline

03

Transformation

Setting up [dbt](#)

Transformation

04

Reporting

Connecting [Looker](#)

01 Storage

Set up BigQuery

Google Cloud

AdventureWorks

Search (/) for resources, docs, products, and more

Search

DISMISS

UPGRADE

Explorer

Search BigQuery resources

Viewing resources.

SHOW STARRED ONLY

adventureworks-431609

SUMMARY

Nothing currently selected

Welcome to BigQuery Studio!

Create new

SQL QUERY

PYTHON NOTEBOOK

DATA CANVAS

Add your own data

Local file

Upload a local file

LAUNCH THIS GUIDE

Google Drive

Google storage service

LAUNCH THIS GUIDE

Google Cloud Storage

Google object storage service

LAUNCH THIS GUIDE

☒ Show welcome page on startup

Job history

REFRESH

02

Ingestion

Setting up Apache Airflow

- [Airflow Documentation](#)
- [Production Deployment Documentation](#)

Writing ELT Python Script

- [.py Code - Extract & Load](#)

```
# importing libraries
```

```
from airflow.decorators import dag, task
from datetime import datetime, timedelta
import requests
import xmltodict
import pandas as pd
import psycopg2
from io import StringIO
```

02

Ingestion

Setting up Apache Airflow

- [Airflow Documentation](#)
- [Production Deployment Documentation](#)

Writing ELT Python Script

- [.py Code - Extract & Load](#)

instantiating DAG

```
@dag(
    schedule=timedelta(minutes=30),
    start_date=datetime(2024, 7, 29),
    catchup=False,
    tags=['Team B']
)
```


02 Writing ELT Python Script

Get Table Lists

```
Sql = """SELECT table_name
        FROM information_schema.tables
        WHERE table_type = 'BASE TABLE'
        AND table_catalog =
'adventure_works'
        AND table_schema
        NOT IN ('pg_catalog',
'information_schema');"""
```

```
@task()
def gt_tbls():
    conn = psycopg2.connect(
        database = "job_listing",
        user = "postgres",
        host= 'localhost',
        password = "password",
        Port = 5432)

    Cursor.execute(sql)
    tbls=cursor.fetchall()
    cursor = conn.cursor()
    conn.commit()
    cursor.close()
    conn.close()
    Return tbls
```



DAG: myjobmag_etl_pipeline

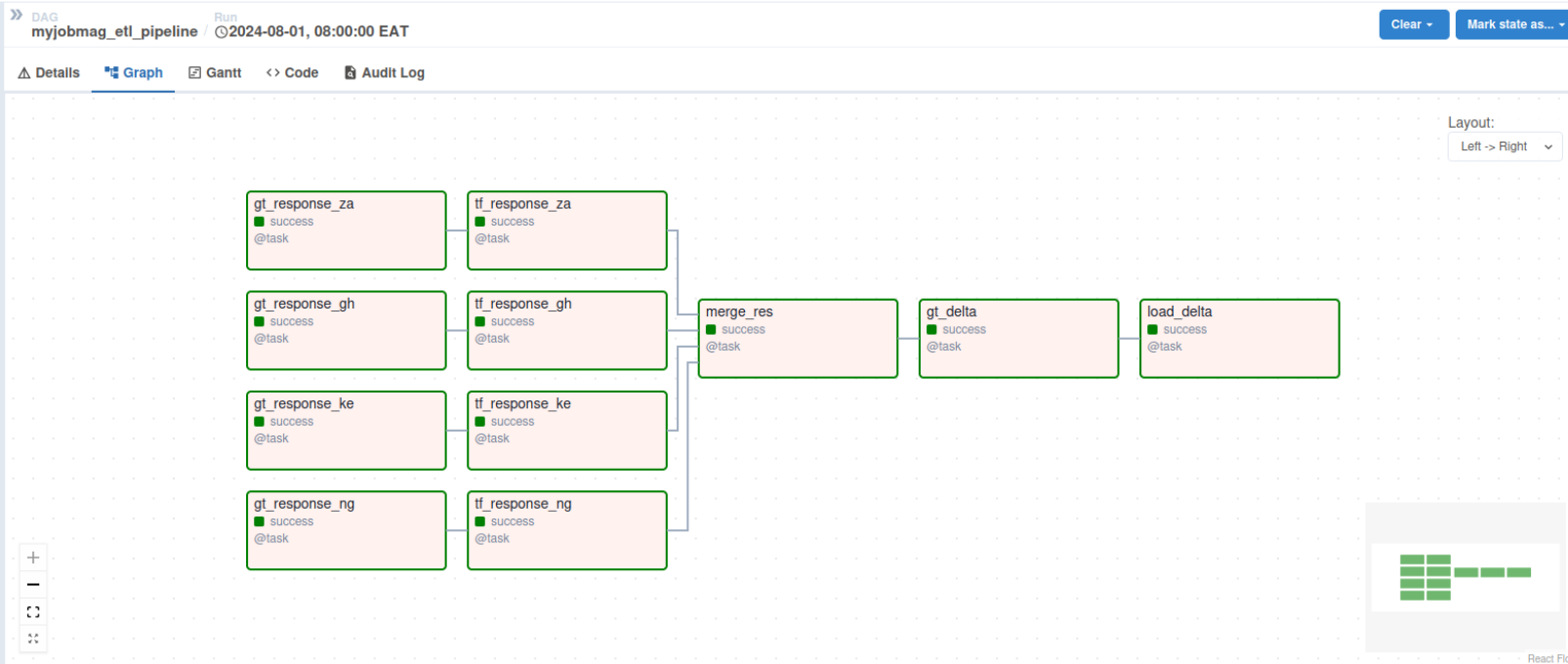
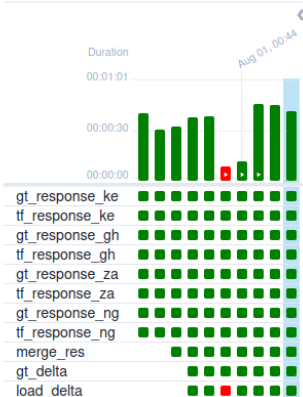
Schedule: 1:00:00 Next Run ID: 2024-08-01, 08:00:00 EAT

08 / 01 / 2024 08 : 31 : 58 AM All Run Types All Run States Clear Filters

Auto-refresh 25

Press **shift** + **/** for Shortcuts

deferred failed queued removed restarting running scheduled shutdown skipped success up_for_reschedule up_for_retry upstream_failed no_status



Jul 30 15:11

pgAdmin 4

FileObjectToolsHelp

Object Explorer

Catalogs

Event Triggers

Extensions

Foreign Data Wrappers

Languages

Publications

Schemas (1)

public

Aggregates

Collations

Domains

FTS Configurations

FTS Dictionaries

FTS Parsers

FTS Templates

Foreign Tables

Functions

Materialized Views

Operators

Procedures (2)

Sequences

Tables (1)

jb_listing

Columns (16)

Constraints

Indexes

RLS Policies

Rules

Triggers

Trigger Functions

Types

Views

Subscriptions

postgres

Login/Group Roles (16)

airflow_user

pg_checkpoint

pg_create_subscription

pg_database_owner

DashboardStatisticsDependentsProcesseslisting_db/postgres...listing_db/postgres...listing_db/postgres...listing_db/postgres@localhost*

listing_db/postgres@localhost

No limit

Data OutputMessagesNotifications

	jid [PK] Integer	pubdate date	expirydate date	jp text	company text	experience text	studies text	Industry text
1	765442	2024-07-30	2024-08-05	Intern VOIP Engineer	Dial Afrika	- years	BA/BSc/HND	ICT / Telecon
2	765434	2024-07-30	2024-08-31	Enterprise-Wide Risk Management Officer	Trade and Development Bank (TDB)	5 - 8 years	BA/BSc/HND , MBA/MSc/MA	Banking / Fin
3	765417	2024-07-30	[null]	Project Finance Administrator	PATH	5 - years	BA/BSc/HND , MBA/MSc/MA	Healthcare /
4	765416	2024-07-30	[null]	Legal Manager - Commercial	Equity Bank Kenya	8 - 10 years	BA/BSc/HND , Diploma	Banking / Fin
5	765415	2024-07-30	[null]	Administrative Assistant, Tatu Girls School	Nova Pioneer	2 - 3 years	BA/BSc/HND , Diploma	Education / T
6	765414	2024-07-30	[null]	History/ CRE Teacher (Eldoret Girls School)	Nova Pioneer	3 - years	BA/BSc/HND	Education / T
7	765413	2024-07-30	2024-08-13	Compliance Officer	African Wildlife Foundation	5 - years	BA/BSc/HND	Travel and Tc
8	765412	2024-07-30	2024-08-05	Pharmaceutical Technologist - Juja	Equity Afia	3 - years	Diploma	Healthcare /
9	765411	2024-07-30	2024-08-05	Pharmaceutical Technologist - Ruiru	Equity Afia	3 - years	Diploma	Healthcare /
10	765408	2024-07-30	2024-08-06	SNE Assistant Officer	Finn Church Aid (FCA)	4 - years	BA/BSc/HND , Diploma	NGO / Non-P
11	765406	2024-07-30	2024-08-12	Private Sector Officer	International Organization for Migration (IOM)	5 - years	BA/BSc/HND , MBA/MSc/MA	NGO / Non-P
12	765389	2024-07-30	2024-08-13	Internship Programme	Turkana County Government	- years	BA/BSc/HND , Diploma	Government
13	765388	2024-07-30	2024-08-11	Sales & Marketing Executive	InspiraFarms	3 - years	BA/BSc/HND	Agriculture /
14	765383	2024-07-30	[null]	Community Service Assistant	Evidence Action	1 - 2 years	BA/BSc/HND , KCSE	NGO / Non-P
15	765379	2024-07-30	2024-08-02	Business Manager	MAL Consultancy	10 - years	BA/BSc/HND , MBA/MSc/MA	Consulting
16	765364	2024-07-30	2024-08-13	Sustainable Livelihoods Officer	Shining Hope For Communities	5 - years	BA/BSc/HND	Consulting
17	765363	2024-07-30	[null]	Learning & Leadership Development Manager	CARE	3 - 5 years	BA/BSc/HND , MBA/MSc/MA	NGO / Non-P
18	765357	2024-07-30	2024-08-15	Key Account Sales Executive	M-Paya	- years	BA/BSc/HND	ICT / Telecon
19	765346	2024-07-30	2024-08-06	Graduate Assistant	Strathmore Business School	- years	BA/BSc/HND	Education / T
20	765342	2024-07-30	[null]	CHP Officer	International Rescue Committee	3 - 6 years	BA/BSc/HND , Diploma	NGO / Non-P
21	765341	2024-07-30	[null]	Finance Officer	International Rescue Committee	3 - years	BA/BSc/HND	NGO / Non-P
22	765326	2024-07-30	[null]	Formulator	ADM	7 - years	BA/BSc/HND	Agriculture /
23	765313	2024-07-30	2024-08-12	Deputy Regional Health Assessment Programme Coordinator	International Organization for Migration (IOM)	7 - 9 years	BA/BSc/HND	NGO / Non-P
24	765303	2024-07-30	2024-08-02	Business Development Manager - Bancassurance	Kingdom Bank Limited	5 - years	BA/BSc/HND	Banking / Fin
25	765296	2024-07-30	2024-08-10	Medical Officer Intern	Avenue Healthcare	- years	BA/BSc/HND	Healthcare /
26	765292	2024-07-30	[null]	Business Development Leader	Visa	12 - years	BA/BSc/HND , MBA/MSc/MA , PhD/Fellowship	Banking / Fin
27	765288	2024-07-30	2024-08-12	Resource Mobilizer	International Transformation Foundation (ITF)	- years	BA/BSc/HND , Diploma	NGO / Non-P
28	765283	2024-07-30	[null]	HR Officer - Performance & Development	Frank Management Consult Ltd	5 - years	BA/BSc/HND	Consulting
29	765282	2024-07-30	[null]	Talent Acquisition Officer	Frank Management Consult Ltd	5 - years	BA/BSc/HND	Consulting
30	765280	2024-07-30	2024-08-16	Human Resources Manager	Rural Agency for Community Development and Assistance (RACID)	5 - years	BA/BSc/HND , MBA/MSc/MA	NGO / Non-P

Total rows: 175 of 175Query complete 00:00:00.074Ln 1, Col 154

