# HoxHunt Web Scrape Pipeline

**Data Sources**

- HoxHunt 2025 Phishing trends report

**Ingestion**

- Python script (HoxHunt_Web_Scrape_Extract_Load_Raw.py)
- GitHub Actions (yearly scheduled pipeline)

**Processing**

Extract:
- Web scrape report
Load:
- Insert raw scraped data into raw.hoxhunt_webscrape
Transform:
- Classify methods into standard phishing categories

**Storage**

PostgreSQL:
- Raw layer: initial webscraped data
- Query tables using HoxHunt_Web_Scrape_SQL_Analysis.ipynb

**Users**

- Dashboards and reports
- Visualize queries in Looker Studio
- Executive Reports