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: inital webscraped data
- Staging layer: staging schema tables
- Warehouse layers

→

Users

- Dashboards and reports
- Security awareness analytics
- Executive Reports