

Ethan Knox

215-703-7885

Philadelphia PA (EST)

ethan@knox.dev



[Github](#)



[Medium.com](#)



[pirate.baby](#)

Containerize everything.

Experience

Hyphen.ai - *Founding team, AI/ML*

September 2023 - April 2024 - Remote contract

Hyphen manages tech teams and infrastructure for clients using AI manager agents.

I built out the entirety of the ML infrastructure, the Generative applications, the prompt design and tuning, the RAG design, and training observability. OSS models only (no OpenAI).

VLLM - SkyPilot - Fastapi - Minstral/Mixtral variants - Langfuse - MemGPT - PGVector - Together.ai - LangChain

- Our production inference intent detection service (my baby) hit a benchmark accuracy of over 82% against our ground truth set and mean latency of 900ms, in less than 2 weeks.
- I built the distributed pipeline architecture for collecting and tagging massive (100M+) tagged infrastructure codebase datasets for classification model training.

Staticdata.io - *Founder*

2021 - 2023 - Remote

[Staticdata.io](#) was a SaaS EL tool that allows users to CRUD Data Warehouse records in an easy-to-use interface. It allowed business users to manage sources of truth for critical reports, goals, KPIs, etc without involving anyone from the data team.

AWS Lambda (SAM), SQS, ECS, Batch, Code Pipeline - Django - Airbyte - Metabase - DBT - BigQuery - Snowflake - Redshift - Azure Data Warehouse/Data Fabric - Terraform

- I designed and wrote all the foundational code, infrastructure, and internal analytics.
- I found our first 10 customers and led our capital raise.

Allied World, a Fairfax Company - AVP Data Architecture

2023 - Remote

Allied World provides insurance and reinsurance solutions to a global market.

I led the design implementation for an international migration to cloud data warehousing, moving from distributed on-prem server farms.

AWS Redshift, Glue, ECS, Lambda (Chalice) - Airflow - Azure SQL Server

- My “innovation” team created best-practices end-to-end analytics stacks and modules to be used by localized implementation teams in multiple APAC, EU, USA and CA zones.
- I designed and my team implemented an unstructured data catalog management system, enabling the IT team to service cataloging across the organization in a single application.

Palmetto Energy - Manager, Data and Analytics

2021 - 2022 - Remote

Palmetto is the leading clean energy marketplace platform in the United States.

Haystack - Snowflake - Looker - Fivetran - DBT - Snowplow Analytics - Tableau - PalmCLI - GCP Compser, K8s, Vertex AI - Elementary - Spectacles - Airflow - Terraform

- I recruited and onboarded a 9-person data team in 6 weeks.
- My team launched an ML predictive analytics suite for determining installation targets, allowing install teams to optimize for just-in-time planning
- A company-wide data roadmap was critical (Palmetto had a dozen siloed data verticals) - I moved them to a single roadmap and an 18-month transition to a federated model
- My team made passing a SOC 2 certification possible
- My team onboarded and trained 150 active Looker users in < 3 months
- My team released PalmCLI Open Source Software, a CLI tool for managing containerized environments

Health Union - *Lead Data Engineer*

2019 - 2021 - Philadelphia PA

Health Union operates a suite of condition-related online communities.

Tensorflow - Snowflake - AWS ECS - Django - Looker - Redash - Airflow - Terraform

- I created the [SnowShu](#) and [XDB](#) OSS libraries, enabling us to move local testing out of Snowflake for a 30% reduction in credit use.
- I moved the data warehouse from a 3NF architecture to a more suitable Kimball warehouse.
- My team migrated the BI tooling from an overworked and failing Redash to a hub-and-spoke Looker solution.
- My team and I designed and built a fully containerized EL & Data Science development-to-deployment ecosystem. This system builds locally in obfuscated containers, ships to a Django-based management tool, and deploys to AWS ECS. Under the hood is an Airflow scheduler. Analysts and Data Scientists can deploy working Jupyter, raw Python, and SQL models into an infinitely scalable environment with a few clicks in a GUI and zero involvement from Engineering or DevOps.

The original estimate for the project was 24 months. With an augmented staff we delivered in 5.

Degreed - *Data Engineer*

2019 - Remote

Degreed is a lifelong learning platform for professional and personal growth.

Azure Data Warehouse, Batch - Airflow - DBT - Domo

- I migrated legacy stored-proc transforms in Azure Data Warehouse to use dbt. To get this working, I created the ADW-DBT OSS adapter port.
- I built a multi-os GUI-based (TKinter) platform for authoring data models (supporting Python and SQL) and deploying to Azure Batch. Job scheduling was done in Airflow.
- I set up and deployed EL processes parallel across the data warehouse and direct support of Domo.

IntegriChain - *Senior Data Warehouse Engineer*

2018-2019 - Philadelphia PA

IntegriChain provides deep-dive analytics for the pharmaceutical industry.

AWS Glue, Athena, Redshift - Infobright

- My team released and maintains [s3parg](#), and OSS library for managing parquet data in S3.
- I architected an immutable hive-based data lake in S3 (accessed via Athena/Redshift).
- I developed and my team deployed a Netflix-style Jupyter Notebook deployment system via Airflow+Papermill, allowing Data Scientists with no software backgrounds to deploy and re-deploy models in production.
- I migrated legacy in-memory Perl transforms to a data-at-rest strategy.

Comoto Holdings Inc. - Data Engineer

2014- 2018 - Philadelphia PA

Comoto Holdings is the parent of motorcycle industry giant brands RevZilla.com, Cycle Gear, and J&P Cycles.

I designed and built the entire data and analytics stack at RevZilla, then integrated the analytics stacks of each merged property as it was acquired.

Snowflake - DBT - Fivetran - Stitch - Looker - Qlik

- I authored and released the [Kimball dbt](#) OSS package.
- I designed and implemented the Snowflake + dbt Kimball Data Warehouse.
- I set up and deployed parallel implementations of Looker and Qlik.

Education

Boston University

Boston MA

Master's in CIS, concentration. Data Analytics

Drexel University

Philadelphia PA

I completed undergraduate classes in network security and Linux administration before pursuing my Master's' at BU.

State University College at Oneonta -Bachelor of Science

Oneonta NY