David Isaac Belais

505 SE 35th Ave, Portland OR 97214 | 503-267-0942 | david@belais.me | david.belais.me | github.com/davebelais

Summary

I am a highly productive software and data engineer with 18 years of experience, including 15 years working in complex, matrixed, enterprise organizations. I pride myself in authoring elegant, bulletproof, type-annotated, well-formed, thoroughly tested, distributable Python libraries and applications, and designing efficient, maintainable, systems which hold up over time. I have extensive experience with Spark and Databricks, myriad analytics and relational databases, and in building rock-solid asynchronous web APIs.

Experience

Nike | Lead Data Engineer, Sustainability Analytics | March 2021 - June 2026

Data Products:

- Nike Sustainability Index (Databricks Lakehouse, Snowflake, Spark, Python, FastAPI, SQLAlchemy, Apache Kafka, AWS Lambda, AWS Aurora PostgreSQL): As the lead data engineer for Sustainable Products and Materials, I worked with material scientists and Sustainability professionals to design and implement the current generation of systems we use to model our scope 3 environmental impacts, including greenhouse gas emissions. We leverage data from Nike's supply chain (such as bills of material, demand planning projections, and purchase orders) to produce scientifically supported, traceable, reproducible, estimates of Nike's environmental impact metrics (such as Kg of CO2-equivalent greenhouse gas emissions). This data was leveraged in the aggregate for annual reporting to stockholders, in more granular reporting to aid product creation in identifying opportunities for improvement. Additionally, we stood up micro-services (FastAPI rocks!) to provide our partners in product creation the means to perform scenario modeling during the material selection and product design process, including identification of functionally equivalent alternate materials to reduce Nike's impact. Because the lifecycle analysis data used to calculate environmental footprint metrics needs updated quarterly, and because footprints needed frequent recalculation for millions of products and hundreds of thousands of unique materials—we employed an atypical design which encapsulated these contributing 3rd party-sourced metrics, along with mapping inputs submitted by business users, in a versioned distributable. This represented a dramatic performance boost over acessing a centralized store for the semi-static data, and could be deployed as a dependency for both Databricks jobs and microservices, ensuring identical output despite not accessing a physical single source of truth.
- Enablon EHS Reporting (Databricks Lakehouse, Snowflake, Spark, Python, SQLAlchemy, OData): I designed and built out the initial implementation of our foundational data ingestion pipelines and warehouse mirroring Nike's 3rd party EHS (Environmental, Health, and Safety) tool used for reporting by our manufacturing partners. This data product was recently handed off to a newly stood-up sister squad to maintain.

Tooling and Frameworks (Python Libraries and CLIs, Github Actions):

- I authored command line tools and frameworks used in CI/CD and locally, numerous SDKs for internal and third-party platforms, extended SQLAlchemy and Alembic to facilitate use of ORMs (object relational mappings) across multiple dialects simultaneously, to facilitate common and complex data frame operations in Spark, to validate data products based on ORM metadata, securely retrieve managed credentials, and many other common and specialized development tasks.

BICP @ Nike | Lead Data Engineer, Sustainability Analytics | March 2020 - March 2021

- Developed a SQLAlchemy-ORM-based framework for automating deployment and versioning (schema migration) supporting all database dialects leveraged by the Nike Enterprise Data & Analytics organization: Databricks, Snowflake, Hive/Presto on S3, and PostgreSQL with full rollback and versioning support
- Authored a framework for Sustainability Analytics' ETL jobs incorporating end-to-end schema-based data validations, local testing, and environment + file system abstraction

BICP @ Nike | Senior Data Engineer, Sustainability Analytics | January 2020 - March 2020

The Kroger Co. | Lead Data Engineer - Web & Digital Analytics | May 2018 - November 2019

- Lead development of data products distilling and exposing analytics to buyers and planners correlating digital and store sales and EBITDA with inventory,sell-through, prices, and promotional events—contributing to decisions resulting in a 56% increase in e-commerce sales in 2018 vs 2017, and a 67% increase in ecommerce sales in 2019 vs 2018

Note: the header above conveys a genericized role name based on current industry usage, for clarity. My official title was "Senior Manager, Web & Digital Analytics".

The Kroger Co. | Lead Data Engineer - Product Information Management | November 2013 - May 2018

- Lead development of multi-platform (Spark/Hive/Presto, Netezza, DB2, Python, SQLAlchemy) OLAP and OLTP data products to ingest, consolidate and normalize sales, dimensional, and click-stream data from disparate subsidiary and partner systems' transactional databases, streaming platforms, APIs, and mainframes
- Engineered algorithms for scoring semi-structured data and performing human-in-the loop data validation and auditing for product descriptions, specifications and photography acquired through trading partners (Python)
- Established source-management capabilities for inbound data to handle complex retailer/vendor/manufacturer relationships (Python)
- Collaborated with emerging digital initiatives to ensure the capture of all metrics needed to facilitate accountability and continuous operational improvement

Note: the header above conveys a genericized role name based on current industry usage, for clarity. My official title was "Manager, Web & Digital Content".

The Kroger Co. (Fred Meyer Stores Inc.) | Systems Analyst, Ecommerce | March 2011 - November 2013

- Researched, designed, and prototyped Fred Meyer's (and later Kroger's) product information management system for customer-facing digital initiatives
- Collaborated with Fred Meyer's technology partner, 1WorldSync, to establish a roadmap, data model, and procedures for sourcing and validating product data from GDSN data pools for use in digital sales channels

Dissent Graphics Inc. | Full-Stack Developer | January 2008 - March 2011

- Designed and developed web applications for clients including: The Garrigan Lyman Group, Microsoft, Best Buy, Avenue A Razorfish, Nereus Communications, BlackEyedPeas.com, TeeFury.com, the Travel Channel's Man v. Food, TheWho.com, Custom Rights, Hello Minor, ExoticTravelers.com, and the ACLU of Oregon

Please see my Linked-In profile for a comprehensive employment history: linkedin.com/in/davidbelais

Education

- Portland State University | Computer Science (Postbaccalaureate) | 2018
- The Art Institute of Portland | Computer Generated Imaging | Bachelor of Science | 2007
- Portland State University | Web Design | 2002 - 2004
- Loyola Marymount University | Fine Arts | 2000 - 2001

Open Source Projects

...because code examples are worth a thousand interview questions!

- dependence: A CLI and library for aligning a python projects' declared dependencies with the package versions installed in the environment in which dependence is executed, and for "freezing" recursively resolved package dependencies (like pip freeze, but for a package, instead of the entire environment).
- maya-zen-tools: An Autodesk Maya extension providing modeling tools for manipulating a polygon mesh using dynamically created NURBS curves and surfaces to distribute vertices and/or UVs
- oapi: A python library for generating client SDKs from Open API documents
- git-author-stats: A CLI and library for extracting periodic author "stats" (insertions and deletions) for a Git repository or Github organization
- gittable: A CLI and library for performing common, but complex, development and CI/CD tasks for a Git repository, such as tagging a commit with your current project/package version and downloading or accessing specific file(s) from a remote repository (including non-public repos)

Please see Github for a full accounting of my activities :-)

Certifications

- AWS Certified Big Data - Specialty
- AWS Certified Cloud Practitioner