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 data engineer and software engineer with 18 years of relevant experience.

I pride myself in:

- Creating resilient, maintainable, integrous data pipelines and products (ETL/ELT batch, micro-batch, and streaming for both OLAP and OLTP databases/lake houses/data lakes/data warehouses) for "big data", master data, fact data and dimensional data
- Authoring elegant, bulletproof, type-annotated, well-formed, thoroughly tested, distributable Python libraries, CLIs, asynchronous micro-services (web APIs), and SDKs
- Designing efficient, maintainable, testable, continuously integrated and deployed, modern software systems (CI/CD, Test-driven development)
- Planning development work with clarity, flexibility, parallel execution, and collaboration in mind (whether using Agile or waterfall)
- Leading engineering teams with complex and ambiguous directives towards clear, executable road maps
- Condensing fact from the vapor of nuance while maintaining traceability and continuity with data provenance, and facilitating data governance
- Utilizing "AI" and machine learning tools and frameworks *selectively*, *prudently*, and with traceability and long-term *cost efficacy* in mind
- Infrastructure as Code: Terraform

Skills

I have professional experience with (not exhaustive):

- Platforms: Databricks, Snowflake, Amazon Web Services
- Languages: Python, SQL, C++, Javascript, HTML, XML, PHP, WSDL, Rust (AWS - including Lambda, EMR, Aurora, IAM, Cloudformation, EC2, S3)
- OLTP Databases: PostgreSQL, MySQL, SQL Server, Oracle, IBM DB2, SQLite, MariaDB
- OLAP "Databases" and query engines: Databricks Lakehouse, Deltalake, Snowflake, Terradata, Netezza, Hive, Presto, DuckDB
- Applications, Services and Frameworks: Apache Spark, Apache Kafka, SQLAlchemy, FastAPI, Flask, Docker, Terraform, Linux, Unix, Github Actions, Jenkins, Kubernetes, Hadoop, Copilot
- Protocols and Specifications: Open API (Swagger), SOAP, MIME, AS2 (for GDSN data pools), ASGI, WSGI
- Distributed File Systems: DBFS, S3, HDFS

Experience

Nike | Lead Data/Software Engineer - Sustainability Analytics | March 2021 - June 2025

Platforms: Databricks, Snowflake, Amazon Web Services (AWS - including Lambda, EMR, Aurora, IAM, Cloudformation, EC2, S3)
Languages: Python, SQL, Javascript
OLTP Databases: PostgreSQL, Oracle, SQLite
OLAP "Databases" and query engines Databricks Lakehouse, Deltalake, Snowflake, Terradata, Hive, Presto
Applications, Services and Frameworks: Apache Spark, Apache Kafka, SQLAlchemy, FastAPI, Docker, Terraform, Linux, Github Actions, Jenkins, Hadoop, Copilot
Protocols and Specifications: Open API (Swagger), ASGI
Distributed File Systems: DBFS, S3, HDFS

- I lead engineering of the Nike Product & Materials Sustainability Index with data products published in Nike's "Sole" Databricks Platform (Lakehouse/Deltalake/Unity Catalog), Snowflake, and (prior to 2023) S3 + hive/presto + EMR. We utilized Spark, Python (applications and libraries, distributed through JFrog Artifactory), ASGI Microservices (Python + FastAPI + SQLAlchemy + Alembic), Apache Kafka (as a publisher), and AWS Aurora PostgreSQL. CI/CD using Jenkins and Github Actions, Terraform for Infrastructure as Code. This entailed design and implementation of systems for distilling material manufacturing process lifecycle assessment data, the expertise of the material scientists and sustainability professionals with whom we collaborated, and materials data from our product creation systems into data products attributing environmental impact measures (greenhouse gas emissions and water quality degradation/depletion) to Nike materials. We subsequently parsed product bills of material line items in order to infer material gross usage (way more involved than it sounds) in kilograms. Applying these measures to purchase order and demand planning data we were able to measure and track aggregate impacts for the enterprise. More importantly, we provided tools for product developers to reduce environmental impacts through better materials selection during the design process. To this end, we stood up micro-services (Python + FastAPI on AWS lambda) for product and material footprint scenario modeling. - I lead development of foundational data products exposing Environmental Health & Safety data from our 3rd-party EHS reporting system, Enablon (via their "Blink" OData API), in Databricks Lakehouse (Unity Catalog), Snowflake, and (prior to 2023) S3 + hive/presto + EMR. I used Spark and Python. - I authored enterprise CLIs (command line interfaces) and frameworks (Python libraries) for use in CI/CD and locally, numerous SDKs (software development kits) for internal and third-party platforms, extended SQLAlchemy and Alembic to facilitate use of ORMs (object relational mappings) across multiple dialects simultaneously, and to facilitate common and complex data frame operations in Spark, validate data products based on ORM metadata, securely retrieve managed credentials, and many other common development tasks.

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

Platforms: Snowflake, Amazon Web Services (AWS - including Lambda, EMR, Aurora, IAM, Cloudformation, EC2, S3)
Languages: Python, SQL, Rust
OLTP Databases: PostgreSQL, Oracle, SQLite
OLAP "Databases" and query engines: Snowflake, Terradata, Hive, Presto
Applications, Services and Frameworks: Apache Spark, SQLAlchemy, Docker, Terraform, Linux, Github Actions, Jenkins, Hadoop
Protocols and Specifications: Open API (Swagger)
Distributed File Systems: S3, HDFS

- I 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.
- I 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/Software Engineer - Sustainability Analytics | January 2020 - March 2020

Platforms: Snowflake, Amazon Web Services (AWS - including Lambda, EMR, Aurora, IAM, Cloudformation, EC2, S3)
Languages: Python, SQL, Rust
OLTP Databases: PostgreSQL, Oracle, SQLite **OLAP "Databases" and query engines:** Snowflake, Terradata, Hive, Presto
Applications, Services and Frameworks: Apache Spark, SQLAlchemy, Docker, Terraform, Linux, Jenkins, Hadoop
Protocols and Specifications: Open API (Swagger), ASGI
Distributed File Systems: S3, HDFS

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

Languages: Python, SQL, Javascript, HTML, XML, WSDL
OLTP Databases: SQL Server, IBM DB2, SQLite **OLAP "Databases" and query engines:** Netezza, Hive, Presto
Applications, Services and Frameworks: SQLAlchemy, Flask, Hadoop, Magento Commerce, IBM Websphere Commerce
Protocols and Specifications: Open API (Swagger), SOAP, MIME, AS2 (for GDSN data pools), WSGI

- I 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.
- I lead development of pricing/promotions and product information integration services for Magento Commerce.