

Owen Sanders

Data-Focused Software Engineer
(Number Plumber)

New York, NY

@helloimowen on github

Skills

Highly skilled in Python, PySpark, SQL, HCL.

Previous experience with C, C++, Java, JS.

Learning Rust, GDScript (Godot game engine), Scala.

Big data tools - Spark ecosystem, databases (OLTP, OLAP, K/V, NoSQL). Kafka. Airflow.

Extensive experience managing Spark clusters using Databricks on AWS.

DevOps Tools - Terraform, Kubernetes, Github Actions, other CI/CD systems.

Experience with EC2, S3, RDS, EMR, Glue, Lambda and many other AWS services.

Experience

Senior Software Engineer, Attentive Inc., New York, New York — Mar 2024 - Present

Data lake platform team. Just got here.

Senior Data Engineer, Tatari Inc., New York, New York — Feb 2023 - Mar 2024

Data Platform team. Led onboarding, mentoring and training for new hires and interns. Led Data Quality Initiative, Metadata Platform Rollout, local clickhouse tooling. Built tooling to improve data quality and observability across the organization

Data Engineer, Tatari Inc., New York, New York — Aug 2020 - Feb 2023

Data Platform team. Owned core data pipeline and infrastructure projects. Worked with Spark, Apache Airflow, Kubernetes, Kafka, Spark, AWS Redshift, S3.

Software Engineer, Blink Health, New York, New York — Nov 2018 - Jun 2020

Data Team - worked to create and maintain an internal data stack.

Worked with sensitive healthcare data in a HIPAA compliant environment.

Projects

Databricks Cost Reduction. Worked across teams to reduce data scanned and Spark job runtime. Reduced data scanned by multiple jobs by more than 95%. Created single node policy in Terraform, a full 1/6th the cost of the previous smallest instance.

DataHub. Led project from PoC through adoption. Integrated and managed Tatari's largest service deployment. Kafka, ElasticSearch on Kubernetes. Postgres on RDS.

Tatari Data Quality. Created Internal data quality tool in Python, based on Great Expectations. Low-code interface capable of running data quality checks against data that lives in the datalake on S3, PostgreSQL, redshift. Owned project from exploration to completion, leading other engineers and onboarding entire engineering organization.

Kafka CDC. Created a Change Data Capture system for Tatari's primary backend Postgres database to stream table changes into the data lake. Used Strimzi Kafka operator for Kubernetes. Transformed data streamed as JSON into parquet using Spark.

Redshift Overhaul. Removed all manual access to Tatari's production redshift cluster. Worked across teams and functions to build replacements for any disparate use cases. Improved cluster performance, organization and administration.

Education

Champlain College, Burlington VT

Bachelor of Science Degree. Major in Computer Science and Innovation.

Specialization in Software Engineering. Minor in Game Programming.

Graduated Cum Laude - May 2018