

Aruã de Carvalho

Senior DevSecOps, Data Platform Engineer

SUMMARY

- Data Platform Engineer and DevOps with 4+ years optimizing data infrastructure and distributed systems using Apache Spark, ClickHouse, and cloud platforms.
- Reduced infrastructure costs by R\$3M+ annually through Spark ingestion framework optimization at XP Inc.
- Expertise in AWS, Terraform, Kubernetes, GitOps, and CI/CD pipelines enhancing platform reliability and scalability.
- Developed observability and event-driven architectures with CloudWatch, OpenTelemetry, Kinesis, and Lambda for real-time workloads.
- Proficient in Python, SQL, TypeScript, and distributed system design, delivering platform tooling adopted by 100+ engineers.

TECHNICAL SKILLS

Main Technical Skills	Apache Spark, Databricks, Apache Airflow, Clickhouse, AWS
Programming Languages	Python, TypeScript
Java Frameworks	Apache Spark
Scala Frameworks	Apache Spark
Python Libraries and Tools	PySpark
Data Analysis and Visualization Technologies	Apache Airflow, Apache Hive, Apache Spark, Databricks
Security	IAM
Databases & Management Systems / ORM	Apache Hive, Apache Spark, AWS DynamoDB, Clickhouse, PostgreSQL, SQL
Cloud Platforms, Services & Computing	AWS, Azure
Azure Cloud Services	AKS, Databricks
Amazon Web Services	AWS CloudWatch, AWS DynamoDB, AWS ECS (Amazon Elastic Container Service), AWS Fargate, AWS Kinesis, AWS Lambda, AWS S3, AWS SQS, AWS VPC
Deployment, CI/CD & Administration	Automation deployment CI, CD, CI/CD, GitOps
Methodologies, Paradigms and Patterns	Distributed systems, Solution Architecture Design

Virtualization, Containers and Orchestration	Docker, Kubernetes, Terraform
Version Control	Github Actions
SDK / API and Integrations	OpenTelemetry
Platforms	Unity
QA, Test Automation, Security	Usability tests
Other Technical Skills	Astronomer, Delta lake, Observability, Peer-to-Peer networks

WORK EXPERIENCE

Data Platform Engineer, DevOps - Independent Contractor (Analytical Platform Rework on ClickHouse Cloud)

Duration: May 2026 - Present

Summary:

- Reworked the analytical platform around ClickHouse Cloud on AWS, enhancing infrastructure and change data capture components while optimizing for multi-tenant workloads
- ClickHouse Cloud is a managed cloud data warehouse service designed for real-time analytics with high performance and scalability

Responsibilities:

- Contributed to infrastructure and CDC components of the platform.
- Tuned platform for multi-tenant workloads to improve query performance and reduce costs.
- Built observability and reliability controls using CloudWatch and OpenTelemetry for production workloads, SLAs, SLOs, and critical alerts.
- Developed AWS infrastructure for isolated environments including VPC networking, IAM, load balancing, masked production-data environments, and database-level tenant isolation.
- Established GitOps-based platform delivery with versioned infrastructure and schema changes, pull-request validation, CI/CD, and automated deployments.

Technologies: ClickHouse Cloud, AWS, CloudWatch, OpenTelemetry, VPC, IAM, GitOps, CI/CD

Data Platform Engineer, DevOps - XP Inc. (Spark Ingestion Framework Optimization and Platform Tooling)

Duration: May 2025 - May 2026

Summary:

- Optimized an organization-wide Spark ingestion framework with automatic source profiling to tune partitioning and cluster configurations, reducing infrastructure costs significantly
- XP Inc



- is a leading Brazilian investment platform offering financial services and products

Responsibilities:

- Optimized Spark ingestion framework with automatic source profiling.
- Built internal SDKs and custom Airflow operators used by over 100 engineers.
- Evolved CI/CD on Databricks Asset Bundles with change management, automated testing, and cost allocation.
- Supported Airflow 3 migration with Astronomer through VNet peering and Kubernetes on AKS, maintaining 99.9% availability.
- Contributed to enterprise migration from Hive Metastore to Unity Catalog.

Technologies: Apache Spark, Databricks, Apache Airflow, Astronomer, Kubernetes, AKS, Unity Catalog

DevOps, Data Platform Engineer - Sacflow (AWS Infrastructure Migration and Event-Driven Architecture Expansion)

Duration: February 2024 - May 2025

Summary:

- Migrated Sacflow's core platform infrastructure to AWS using Terraform, establishing isolated environments and controlled infrastructure changes
- Expanded event-driven architecture to support real-time product workloads and operational reporting

Responsibilities:

- Migrated platform infrastructure to AWS with Terraform including VPC networking, IAM, load balancing, and ECS on Fargate.
- Established isolated environments, remote state with locking, controlled infrastructure changes, and plan-gated pull requests.
- Built CI/CD pipelines with GitHub Actions, automated test gates, ECR image versioning, and blue/green deployments on ECS.
- Automated staging and sandbox environments with cross-environment data replication for release validation.
- Expanded event-driven architecture using Kinesis, Lambda, and SQS for real-time workloads and reporting.

Technologies: AWS, Terraform, VPC, IAM, ECS, Fargate, GitHub Actions, Kinesis, Lambda, SQS

Software Engineer - Sacflow (Real-Time Messaging and Search Platform Enhancements)

Duration: February 2023 - January 2024

Summary:

- Improved Sacflow's real-time messaging platform by optimizing WebSocket traffic and implementing full-text search and SLA event processing
- Sacflow is a technology company focused on data-driven product solutions



Responsibilities:

- Reduced WebSocket traffic by replacing broadcast-and-filter with DynamoDB-based connection routing.
- Built real-time full-text search using Typesense across hundreds of millions of messages.
- Developed SLA event processing for response and resolution metrics and real-time messaging performance metrics.
- Redesigned media caching strategy to reduce S3 storage costs.

Technologies: WebSocket, DynamoDB, Typesense, S3

EDUCATION

- **FIAP | Technologist Degree in Data Science and Data Engineering**
2024 - 2026

CERTIFICATION

- **AWS Certified Data Engineer - Associate**
2025
- **Databricks Certified Associate Developer for Apache Spark**

