Brenda Streit Schussler

Data Engineer

Brazil | +55 55 991131885 | brendaschussler@gmail.com

LinkedIn: in/brendaschussler | GitHub: github.com/brendaschussler

Professional Summary

Data Engineer with hands-on experience designing and maintaining ETL/ELT data pipelines on Azure-based data platforms. Experienced with Apache Spark, Databricks, Azure Data Factory, and Apache Kafka, building both batch and streaming pipelines for distributed data processing. Strong background in troubleshooting complex production issues, incident investigation, and root cause analysis (RCA), with a focus on improving performance, data quality, and governance. Analytical mindset, impact-driven approach, and experience working in dynamic, high-demand environments.

Selected Achievements

- Resolved a critical large-scale data duplication issue involving billions of records by performing root cause analysis and implementing idempotent processing, reducing pipeline execution time from 6–8 hours to ~30 minutes.
- Developed end-to-end data pipelines using Apache Kafka, Azure Data Factory, and Databricks for large-scale data ingestion and processing (batch and streaming).
- Designed and implemented pipelines following the Lakehouse (Medallion) architecture, ensuring standardization, data governance, and quality.
- Author of a peer-reviewed scientific paper on an IoT network traffic collection platform, accepted at the International Conference on Network and Service Management.
- Author and co-author of 7 scientific papers in international high-performance computing conferences (CARLA, PDP, EUROPAR, ERAD), including 2 Best Paper Awards.

Technical Skills

Languages & Processing
Python, SQL, PySpark, Apache Spark, Spark SQL

Data & Architecture
Data Lake, Data Warehouse, Lakehouse, Delta Lake, Medallion Architecture (Bronze, Silver, Gold), Data Modeling (Relational & Dimensional – Star Schema, Snowflake)

Pipelines & Streaming
Apache Kafka, ETL, ELT, Spark Structured Streaming, Batch Processing, Stream Processing, Data Ingestion, Data Pipelines

Cloud & Platforms
AKS (Kubernetes), Containerization, Application Insights, Pipeline Monitoring, Logging

Infrastructure & Observability
AKS (Kubernetes), Containerização, Application Insights, Monitoramento de Pipelines, Logging

Other
Data Governance, Data Quality, Data Validation, Data Consistency, Root Cause Analysis (RCA), Troubleshooting, Performance Optimization, Technical Documentation, Git

Professional Experience

Data Engineer | KIS Solutions
Mar 2026 – Present

Data Engineering Intern
Aug 2025 – Feb 2026

Work in data engineering within an Azure-based data platform, building and maintaining pipelines for large-scale data ingestion, processing, and transformation.

- Built end-to-end ETL pipelines using Apache Kafka, including topic creation and configuration, orchestration in Azure Data Factory, processing in Databricks (Apache Spark), and persistence into Delta Lake tables.
- Independently identified and resolved a structural flaw in a production data pipeline causing billion-scale record duplication, leading the end-to-end root cause analysis and implementing an idempotent processing strategy that reduced pipeline runtime from 6–8 hours to 30 minutes while restoring data correctness.
- Presented the issue, technical analysis, and remediation approach to the engineering team.
- Developed Databricks notebooks consuming Kafka topics in both streaming and batch modes, leveraging Spark Structured Streaming for incremental real-time processing.
- Designed and maintained data pipelines using Azure Data Factory and Databricks, supporting ingestion and transformation of structured and unstructured data.
- Implemented Data Lake modeling following the Medallion Architecture (Bronze, Silver, Gold) and applied standardized dataset naming and governance practices.
- Investigated production incidents by analyzing AKS (Kubernetes) pods, Application Insights logs, Databricks jobs, ADF pipelines, and Kafka streams, performing root cause analysis and documenting findings.
- Extensively used Python and SQL across Databricks notebooks, MySQL, and SQL Server-based Data Warehouses.

Mobile App Developer for Data Collection | Concordia University - Montreal, Canada

Mar 2025 – Jul 2025

Developed a mobile platform for collecting, processing, and visualizing IoT network traffic data, supporting research in cybersecurity and connected environments.

- Developed an Android application in Kotlin to capture network traffic packets and device events from connected IoT devices.
- Implemented features for structuring and storing network telemetry data for monitoring and analysis.
- Integrated Firebase authentication, real-time database synchronization, and cloud storage for captured data.
- Built interfaces for real-time visualization of network activity and connected devices.
- The platform is currently used in a research initiative building a large-scale labeled dataset of IoT network traffic from more than 50 devices.

- Authored and published a peer-reviewed paper describing the architecture and implementation of the IoT network traffic data collection platform, accepted at the International Conference on Network and Service Management (CNSM) 2025.

Researcher in Parallel and Distributed Computing and Data-Intensive Systems | Petrobras, UFRGS - Brazil

Aug 2022 – Feb 2025

Worked on high-performance computing (HPC), performance optimization, and data-intensive scientific simulations in heterogeneous computing environments.

- Designed and validated a genetic algorithm for CUDA kernel autotuning, optimizing GPU resource usage and reducing execution time by over 80%.
- Conducted performance and energy-efficiency benchmarks across cloud and local infrastructures to support architectural decisions for scientific workloads.
  Developed cross-platform implementations of the Fletcher method in Python, CUDA, and OpenACC, improving portability and reproducibility.
- Implemented runtime autotuning with persistent configuration storage, enabling infrastructure-aware optimization and achieving 20% average improvement in Energy-Delay Product (EDP).
- Processed and analyzed large volumes of experimental data from scientific simulations, performing exploratory analysis and optimization of execution metrics.
- Documented pipelines and experimental methodologies to ensure reproducibility and research continuity.
- Co-authored 7 peer-reviewed papers in international conferences including CARLA, PDP, EUROPAR, and ERAD, receiving 2 Best Paper Awards and one CAPES Qualis A3 publication.

Education

Bachelor's Degree in Computer Engineering
Federal University of Rio Grande do Sul (UFRGS)
2020 – 2026

International Experience

Mitacs Globalink Research Internship | Concordia University - Montreal, Canada

Selected for a competitive international research program funded by Mitacs, collaborating in a multicultural research environment with an international team.

- Conducted applied research in network data collection and analysis for IoT environments.
- Developed strong cross-cultural collaboration and technical communication skills in an international research setting.

Languages

Portuguese: Native
English: Advanced
Spanish: Intermediate