# **Matthew L**

Senior Data Engineer | Python Engineer | Cloud Data Architect

[mlong@magneton.io](mailto:mlong@magneton.io) | 312-478-9956 | [LinkedIn](http://www.linkedin.com/in/mlabs1913)

# **Executive Summary**

Senior Data Engineer with 20+ years building scalable data pipelines, analytics platforms, and cloud-native data architectures for Fortune 500 companies and high-growth startups. Deep expertise in Python-based data engineering, AWS data services, real-time processing, and AI/ML integration. Proven track record delivering data solutions that drive business intelligence and operational efficiency. Patent holder in distributed systems with extensive experience in ETL/ELT, data warehousing, and streaming architectures.

## **Key Achievements**

* **$300M Exit** - Core contributor at Evident.io (acquired by Palo Alto Networks)
* **$70K Annual Savings** - Optimized AWS data infrastructure reducing operational costs
* **100+ Data Pipeline Implementations** - Delivered high-value solutions across Finance, Healthcare, Energy, and Government
* **Patent Holder** - Invented secure distributed messaging architecture
* **Real-Time Processing** - Built systems handling millions of events daily with sub-second latency
* **500+ Commits (2024-2025)** - Active development across data engineering projects
* **AI-Powered Data Solutions** - Integrated Vision LLMs and OCR for intelligent document processing

# **Core Competencies**

## **Data Engineering Excellence**

* **Data Pipeline Development:** Expert in ETL/ELT, real-time streaming, batch processing
* **Python Engineering:** Advanced Python for data processing, orchestration, and automation
* **Cloud Data Architecture:** AWS Glue, Redshift, Kafka, Kinesis, S3, AppFlow
* **Big Data Technologies:** PySpark, Apache Kafka, distributed processing frameworks
* **Infrastructure as Code:** AWS CDK, Terraform, CloudFormation

## **Technical Expertise**

### **Data Engineering Stack**

* **Languages:** Python (primary), TypeScript, SQL, Scala
* **Data Processing:** PySpark, Pandas, NumPy, Dask
* **Orchestration:** AWS Glue, Step Functions, Airflow
* **Streaming:** Apache Kafka, AWS Kinesis, real-time event processing
* **Data Warehousing:** Redshift, Snowflake, BigQuery
* **Storage:** S3, Delta Lake, Parquet, data lake architecture

### **Cloud & Infrastructure**

* **AWS Data Services:** Glue, Redshift, Kinesis, AppFlow, Lake Formation, Athena
* **DevOps:** Docker, Kubernetes, CI/CD for data pipelines
* **IaC:** AWS CDK, Terraform, infrastructure automation
* **Monitoring:** CloudWatch, Datadog, pipeline observability

### **AI/ML Integration**

* **Vision AI:** AWS Rekognition, Textract, Vision LLMs for document processing
* **ML Pipelines:** SageMaker, TensorFlow, PyTorch integration
* **AI Tools:** Codex, Claude, GPT for code generation and optimization

# **Education & Professional Development**

* **M.S. & B.S. Computer Science** - University of Colorado
* **Certifications:** AWS Solutions Architect, Red Hat Certified
* **Continuous Learning:** Advanced studies in Distributed Systems, AI/ML, and Data Architecture

# **Professional Experience**

## **Pop Secret/PopCorners (2025)**

***Senior Software Engineer***

**Impact:** Created national marketing campaign’s technical infrastructure

* Built Twilio-powered IVR system for “Buttery Secrets Hotline”
* Implemented voice recording, storage, and playback features with S3 data pipeline
* Designed 4-option interactive menu with SMS integration
* Deployed serverless architecture on Vercel with AWS S3 storage
* Created data ingestion pipeline for call analytics and reporting
* **Tech Stack:** Node.js, Express, Twilio, AWS S3, PostgreSQL, Vercel

## **TSI (2024-2025)**

***Senior Full-Stack Engineer***

**Impact:** Architected cloud infrastructure and data-driven applications

* Designed AWS infrastructure with React/Laravel application stack
* Implemented Kubernetes orchestration for scalable deployments
* Built responsive frontend components with modern React patterns
* Optimized database queries improving response times by 40%
* Developed data integration patterns for multi-service architecture
* **Tech Stack:** AWS, React.js, PHP Laravel, Kubernetes, PostgreSQL

## **Zoro/Grainger (2024-2025)**

***Senior Software Engineer***

**Impact:** Built data-intensive microservices for e-commerce platform

* Developed shipment tracking service processing millions of real-time tracking events
* Implemented data pipelines for FedEx webhook integration with event streaming
* Built return service with complex data aggregation and analytics
* Created ETL processes for order data synchronization across systems
* **Tech Stack:** Node.js, Python, Java/Spring Boot, Google Cloud, Kafka, BigQuery

## **Cogability (2023-2025)**

***Senior Data Engineer***

**Impact:** Built intelligent data pipelines for automated municipal ticket processing

* Architected end-to-end data ingestion pipeline processing 100K+ PDF documents monthly
* Integrated AWS Textract OCR with Vision LLMs for intelligent document classification
* Developed Python-based ETL workflows extracting structured data from unstructured PDFs
* Implemented data validation and quality checks ensuring 99.5% accuracy
* Built real-time monitoring dashboards tracking pipeline performance and data quality
* **Tech Stack:** Python, AWS Lambda, Textract, S3, Vision LLMs, DynamoDB, CloudWatch

## **Shell (2022-2023)**

***Senior Software Engineer***

**Impact:** Architected enterprise data analytics platform for energy forecasting

* Designed microservices data architecture using Kafka as event backbone
* Built real-time data processing pipelines reducing forecasting latency by 40%
* Implemented ML-powered predictive analytics on time-series energy data
* Created data lake on AWS S3 with Glue catalog for unified data access
* Developed PySpark jobs for large-scale batch processing
* **Tech Stack:** Python, PySpark, Kafka, AWS EKS, S3, Glue, Redshift

## **PredictiveHR (2021-2023)**

***Senior Python & Data Engineer***

**Impact:** Developed predictive analytics platform for workforce planning and HR forecasting

* Designed and implemented AWS CDK infrastructure for scalable data warehouse solution
* Built PySpark jobs processing millions of HR records for predictive modeling
* Architected Redshift data warehouse with star schema optimized for analytics queries
* Created data integration pipelines using AWS Glue orchestrating ETL workflows
* Implemented AppFlow connectors for SaaS integration (Workday, ADP, BambooHR)
* Developed unit testing framework for Lambda functions achieving 90% code coverage
* Built ML models predicting employee attrition with 85% accuracy
* **Tech Stack:** Python, AWS CDK, PySpark, Redshift, AWS Glue, AppFlow, Lambda, SageMaker

## **BetaCom (2020-2021)**

***Lead Data Engineer***

**Impact:** Architected real-time data ingestion platform for 5G network analytics

* Designed high-throughput data ingestion architecture processing 50M+ events daily
* Built Python FastAPI async services handling concurrent data streams with sub-second latency
* Implemented Kafka-based streaming pipeline for real-time network telemetry
* Developed React dashboard visualizing live network performance metrics
* Optimized data partitioning strategies reducing query times by 60%
* Created data quality monitoring detecting anomalies in real-time
* **Tech Stack:** Python, FastAPI (async), Apache Kafka, AWS MSK, React, PostgreSQL, Redis

## **Fidelity Labs (2020-2023)**

***Senior Software Engineer***

**Impact:** Built data pipelines for RegTech platform processing financial transactions

* Architected multi-cloud data infrastructure supporting 10K+ transactions/second
* Implemented real-time data streaming with Kafka handling regulatory data
* Built ETL pipelines integrating multiple financial data sources
* Developed BERT-based NLP models for document classification
* Created data quality frameworks ensuring compliance with financial regulations
* **Tech Stack:** Python, Java Spring, Kafka, AWS, GCP, BigQuery, Airflow

## **Regis Corporation (2018-2022)**

***Senior Full-Stack Developer***

**Impact:** Developed ML-powered document processing system

* Built data pipeline processing 100K+ financial documents monthly
* Implemented AWS SageMaker ML models for document classification
* Created automated data extraction from PDFs using AWS Textract
* Developed data warehouse for portfolio analytics serving 500+ users
* **Tech Stack:** Python, AWS SageMaker, Textract, Lambda, Redshift, React

## **Evident.io (2016-2018)**

***Senior Software Engineer***

**Company Exit: $300 Million to Palo Alto Networks**

* Built data pipelines processing 1M+ security events daily
* Optimized data storage and processing reducing AWS costs by $70K annually
* Implemented real-time alerting system with sub-second event detection
* **Tech Stack:** Python, Ruby, PostgreSQL, Redis, Elasticsearch

# **Technical Architecture Expertise**

## **Data Engineering Patterns**

* **Data Integration:** CDC, batch/streaming hybrid, event-driven architectures
* **Pipeline Design:** Lambda/Kappa architecture, medallion architecture (bronze/silver/gold)
* **Data Quality:** Great Expectations, Deequ, custom validation frameworks
* **Performance Optimization:** Partitioning strategies, compression, query optimization
* **Testing:** Unit testing for Lambda functions, integration testing for pipelines

## **Operational Excellence**

* **Infrastructure as Code:** AWS CDK for data infrastructure automation
* **Monitoring & Observability:** CloudWatch, Datadog, pipeline health metrics
* **Data Governance:** Cataloging, lineage tracking, access control
* **Cost Optimization:** Resource right-sizing, storage tiering, query optimization
* **Security:** Encryption at rest/transit, IAM policies, compliance frameworks

# **Industry Impact**

**Financial Services:** Fidelity, Avant Credit, First Data, Blue Cross Blue Shield **Energy & Utilities:** Shell, Houston Energy, BetaCom 5G networks **Technology:** Apple, Target, Zoro/Grainger e-commerce **Government:** DARPA, USDA **SaaS:** PredictiveHR, Cogability municipal services

# **Value Proposition**

I bring deep data engineering expertise with a focus on:

* **Scalable Data Pipelines** - Building robust, performant ETL/ELT workflows
* **Python Engineering** - Advanced Python for data processing and automation
* **Cloud-Native Architecture** - Leveraging AWS data services (Glue, Redshift, Kafka)
* **AI/ML Integration** - Incorporating Vision LLMs and ML models into data workflows
* **Infrastructure as Code** - AWS CDK and Terraform for reproducible infrastructure
* **Testing & Quality** - Unit testing Lambda functions and data validation frameworks

# **Ready to Build Data Solutions Together?**

Let’s discuss how my data engineering expertise can help transform your data infrastructure and unlock business value.

**Connect:** [mlong@magneton.io](mailto:mlong@magneton.io) | 312-478-9956 | [LinkedIn](http://www.linkedin.com/in/mlabs1913)

*References and detailed project portfolios available upon request*