

Elijah Perez

Lead Data Engineer/Scientist | Distributed Systems & ML Infrastructure
Albuquerque, NM | [REDACTED] | [REDACTED] | [LinkedIn](#)

Professional Summary

Lead Data Engineer and Data Scientist with a Master's in Applied Mathematics and experience building distributed ML and data infrastructure at enterprise-scale. Supports systems processing ~1 billion log events per day, with direct ownership over components handling millions of daily events across multi-cloud environments (AWS, Azure, GCP).

Leads architecture and delivery of secure, production ML pipelines, cloud infrastructure, and real-time inference workflows, with a focus on observability, system reliability, and operational compliance in mission-critical federal environments. Strong background in backend engineering (Java/Python), distributed systems, and cloud-native architecture, with full ownership of multidisciplinary projects, system reliability, and end-to-end delivery.

Experienced in collaborating across distributed teams and time zones, driving alignment across engineering, platform, and analytics organizations. Holds a Public Trust (MBI) clearance and operates effectively in high security, large-scale environments.

Professional Experience

Lead Data Engineer | LynkBlox (IRS Contractor), Remote | 2025 – Present

Technical Lead (Observability & Telemetry)

- Lead architecture and development of enterprise-scale distributed telemetry and ML driven monitoring systems supporting mission-critical cloud and SaaS platforms, powering large-scale systems that process ~1 billion log events per day.
- Serve as technical lead for 3 of 6 major initiatives, representing ~50% of IRS-wide cloud monitoring efforts across AWS and Azure.
- Design and implement secure, production grade data pipelines and deployment workflows enabling real-time and batch ML inference, with emphasis on reliability, monitoring, and operational consistency in distributed environments.
- Own end-to-end system design for telemetry ingestion, normalization, and serving layers, integrating AWS CloudTrail, CloudWatch, and Config data into high performance datasets.
- Build scalable service modeling infrastructure, automating generation of 100+ service hierarchies, replacing manual configurations and improving system consistency.
- Design and deployed ML based anomaly detection systems integrated into production monitoring pipelines, utilizing statistical baselines, adaptive thresholds, and time series techniques.
- Lead adoption of AIOps systems for event correlation and incident aggregation, significantly reducing alert noise and improving operational response.
- Drive multidisciplinary team initiatives to resolve data quality issues across distributed systems, ensuring high availability and reliability of production data feeds.
- Act as technical lead and system owner, guiding architecture decisions, mentoring engineers, and coordinating across platform, ingestion, and analytics teams while delivering end-to-end monitoring coverage for multi-cloud (AWS, Azure, GCP) environments across the IRS enterprise.

Data Scientist II | Noblis (IRS Contractor), Remote | 2023 – 2025

ML Systems & Infrastructure Focus

- Led development of enterprise-scale ML systems for behavioral analytics and risk scoring, processing large volumes of structured and semi-structured log data.
- Designed and deployed distributed ML pipelines using Spark/Databricks, enabling scalable training and inference across enterprise datasets.
- Built and productionized neural network based models (TensorFlow, PyTorch) integrated into operational workflows for real-time risk evaluation.
- Developed NLP and LLM-based context aware analysis systems to extract structured insights from unstructured log data, improving downstream model performance and signal detection.
- Engineered end-to-end systems supporting data ingestion, feature engineering, model training, and inference, bridging research and production environments.
- Implemented statistical monitoring frameworks (Prophet, IQR, Z-score) to ensure data integrity and reliability of ML driven systems.
- Collaborated with platform teams to optimize data availability, schema consistency, and system performance across distributed pipelines.

Data Scientist I | Noblis (IRS Contractor), Remote | 2022 – 2023

- Designed and deployed machine learning systems for financial risk prediction, including time series forecasting, random forest models, and neural networks.
- Built distributed data processing pipelines (Spark/Databricks) handling structured financial datasets and unstructured JSON logs.
- Developed and optimized SQL/NoSQL queries to support large-scale model training and evaluation workflows.
- Conducted statistical validation and experimentation (A/B testing) to measure model performance and system impact.
- Created dashboards and reporting tools to communicate insights to stakeholders and support decision-making.

Technical Skills & Core Competencies

- **Languages & Backend:** Java, Python (PySpark, Pandas, NumPy), REST APIs, Backend Service Development
- **Distributed Systems & Data:** Spark, Databricks, SQL, ETL/ELT Pipelines, Data Modeling, Schema Normalization, Log Processing, Batch & Distributed Systems
- **Machine Learning Systems:** TensorFlow, PyTorch, Neural Networks, NLP, Time Series (Prophet, ARIMA), Statistical Anomaly Detection, Feature Engineering, ML Inference Workflows
- **Cloud & Infrastructure:** AWS (CloudTrail, CloudWatch, Config), Azure, GCP, Cloud Telemetry Pipelines, Data Reliability Systems
- **Observability & Security:** Splunk Enterprise, ITSI, Enterprise Security, KPI Modeling, Risk-Based Alerting, Event Correlation
- **Performance & Reliability:** Latency Reduction, Throughput Optimization, Data Quality, System Reliability
- **Leadership & Ownership:** Technical Leadership, Cross-Functional Collaboration, End-to-End System Ownership

Education

Master of Science in Applied Mathematics

University of New Mexico (2019 – 2021)

Bachelor of Science in Mathematics, Minor in English

University of New Mexico (2015 – 2019)