SEAN LUKA GIRGIS

Senior Data Engineer | Capacity & Infrastructure Optimization

seanlgirgis.github.io | 214-315-2190 | seanlgirgis@gmail.com | GitHub | LinkedIn | X / Twitter

Senior Data Engineer & Cloud Architect with 20+ years of enterprise experience. Specialized in migrating legacy on-prem pipelines to **Serverless AWS Architectures** (Glue/Athena). Expert in **PySpark**, **GenAI Agents** (Text-to-SQL), and high-scale **Capacity Forecasting**.

FLAGSHIP PROJECTS

Serverless Data Platform (AWS)

- **Architecture**: Designed a Serverless Lakehouse using S3, Glue Catalog, and Athena.
- **AI Agent**: Built a 'Text-to-SQL' GenAI bot using **Claude 3 Sonnet** to democratize data access.
- **ETL Optimization**: Fixed 'small file' issues by implementing **Snappy Parquet** compression.

AWS Glue · Athena · Bedrock (GenAI) · S3 · PySpark

Project: HorizonScale — Modernizing Enterprise Capacity with AI & PySpark

- **Why I built this:** To replace legacy, manual 'Trenda' processes with a modern, agentic data pipeline capable of handling banking-scale telemetry.
- **Turbo Prophet:** Architected parallel generator-based pipeline reducing forecasting cycles by **90%**
- **Visual Analytics:** Built interactive Streamlit dashboard serving real-time capacity insights and 'High Trust' utilization scores.

Python · Prophet · Streamlit · Spark · Multiprocessing

PROFESSIONAL EXPERIENCE

Senior Capacity & Data Engineer   2017 – Dec 2025
CITI

- Architected automated **ETL pipelines** using **Python** and **Pandas** to ingest P95 performance telemetry from 6,000+ endpoints (**BMC TrueSight/TSCO**), replacing manual Trenda processes.
- Designed and optimized **Oracle Database** schemas for historical data retention, enabling long-term trend analysis and seasonal risk forecasting.
- Developed **ML-driven forecasting models** using Prophet and scikit-learn to predict infrastructure bottlenecks 6 months in advance, improving provisioning accuracy.
- Integrated disparate data feeds (CSV, Excel, TSCO) into a unified **Oracle** reporting framework, providing executive dashboards with real-time utilization insights.

Performance Engineer   2017 – 2017
G6 Hospitality LLC & HCL/ENTERGY

- Managed end-to-end monitoring for Brand.com using Dynatrace AppMon and Synthetics.
- Led the "FAST" project, data mining user performance metrics to optimize critical money-generating systems.

Senior Consultant (CA APM Consulting)

CA Technologies & TIAA-CREF

- Led enterprise CA APM implementations and upgrades (v9.1 → v10.1) for financial clients, managing 4,000–6,000 agents across multi-cluster environments.
- Served as CA APM SME, handling daily operations, cluster maintenance, agent/Power Pack rollouts, and performance bottleneck resolution in J2EE/.NET stacks.
- Designed custom dashboards, alerts, reports, and Perl/Ksh data-extraction scripts to deliver actionable insights for business-critical monitoring.
- Provided architectural sizing recommendations, Golden Image creation, client training, and technical team leadership for APM deployments.

EXPERIENCE HISTORY

Sabre

2008 – 2012

Senior Systems & Data Migration Engineer

Sprint/Corpus

2001 – 2008

Developer - High Availability Interfaces (C++)

CORE SKILLS MATRIX

Data Engineering & Cloud

- **Pipeline:** Python (Pandas/Generators), SQL (Oracle/Postgres), PySpark.
- **Cloud:** AWS S3 (Lifecycle Rules), Glue, Athena.
- **Big Data:** Oracle Partitioning, Parquet Optimization.

High-Performance Systems & AI

- **Forecasting:** Prophet, Scikit-learn, Time-Series Modeling.
- **Backend:** C++, OCCI, Java, Multiprocessing, Linux/Shell.
- **Tools:** Airflow, Git, Docker, BMC TrueSight, AppDynamics.

EDUCATION & CERTIFICATIONS

Post-Graduate Diploma, Computer Science

Humber College

Bachelor of Science, Engineering

Zagazig University