

GUNASEKARAN RAVI

DATA ENGINEER | AZURE DATA ENGINEER

Chennai, India | gunasekaranravieng@gmail.com | github.com/Gunasekaranravieng

PROFESSIONAL PROFILE

Data Engineer with 4+ years of experience building and maintaining Microsoft Azure data pipelines. Hands-on experience with Azure Data Factory, Azure Databricks, PySpark, SQL and Delta Lake across batch ingestion, incremental processing, Medallion Architecture, data quality and analytics-ready datasets. Experienced in developing end-to-end ETL/ELT workflows, Spark transformations, Delta Lake processing, pipeline monitoring and Power BI reporting.

TECHNICAL SKILLS

Cloud & Azure	Azure Data Factory, Azure Databricks, Azure Data Lake Storage (ADLS), Azure SQL, Azure Synapse
Data Engineering	ETL/ELT, Batch Processing, Incremental Loads, Medallion Architecture, CDC, Data Quality, Reconciliation
Processing & Lakehouse	Apache Spark, PySpark, Delta Lake, Spark SQL, Spark Structured Streaming
Programming & Querying	Python, SQL
Analytics & Formats	Power BI, Parquet, Delta, CSV, JSON
Engineering Tools	Databricks Jobs, Git, GitHub, GitHub Actions, Pipeline Monitoring, Technical Documentation

PROFESSIONAL EXPERIENCE

Data Engineer | Opseazy Solutions Private Limited **Aug 2022 - Present**

- Designed and maintained Azure Data Factory pipelines for batch ingestion, configuring linked services, datasets, triggers and integration runtimes.
- Implemented Copy, Lookup, ForEach, Web and Notebook activities and developed incremental loads using watermark columns and last-modified timestamps.
- Developed Azure Databricks notebooks using Python and PySpark for cleansing, joins, aggregations, window functions and business-rule transformations.
- Handled schema mismatches, corrupt files and late-arriving data, and optimised Spark workloads using partitioning, caching and efficient transformations.
- Implemented Delta Lake MERGE/UPSERT, OPTIMIZE, VACUUM and schema evolution while maintaining Bronze, Silver and Gold data layers.
- Loaded curated data into Azure SQL, Azure Synapse and ADLS and used SQL for validation, reconciliation and data-quality checks.
- Automated batch workflows with Databricks Jobs, monitored production pipelines, resolved failures within SLA and maintained pipeline/notebook documentation.

PROFESSIONAL PROJECT

LearnSoft Training Institute - Enquiry Data Pipeline | 2022 - Ongoing

Tech: MongoDB, Azure Data Factory, ADLS, Azure Databricks, PySpark, Power BI

- Built an end-to-end pipeline to ingest student-enquiry data from MongoDB into ADLS using Azure Data Factory and transform it in Azure Databricks.
- Configured ADF linked services, datasets and copy activities for semi-structured JSON records.
- Developed PySpark transformations for null handling, deduplication, phone-number standardisation and derived enquiry attributes.
- Prepared curated reporting datasets and Power BI dashboards for enquiry trends, course interest and lead-volume analysis.

SELECTED DATA ENGINEERING PORTFOLIO PROJECTS

Enterprise Sales Lakehouse - Flagship Project

Tech: PySpark, Apache Spark, Delta Lake, Spark SQL, GitHub Actions

- Built a Medallion Architecture workflow from source generation through Bronze, Silver and Gold analytical layers.
- Implemented incremental processing with Delta MERGE, data-quality checks, reconciliation, audit logging, monitoring and business analytics.
- Repository: github.com/Gunasekaranravieng/Enterprise-sales-lakehouse

Real-Time E-Commerce Lakehouse

Tech: PySpark, Apache Spark, Delta Lake, CDC, SCD Type 2

- Implemented Bronze/Silver/Gold processing with CDC incremental upserts, idempotent Delta MERGE and SCD Type 2 historical dimension handling.
- Added data-quality validation, invalid-record quarantine, audit metrics and reconciliation.
- Repository: github.com/Gunasekaranravieng/real-time-ecommerce-lakehouse

Real-Time IoT Streaming Lakehouse

Tech: PySpark, Spark Structured Streaming, Delta Lake, Event-Time Processing

- Built an IoT streaming Lakehouse using event-time processing, watermarking for late events and five-minute window aggregations.
- Implemented anomaly indicators, sensor data-quality rules, reconciliation and pipeline monitoring using demonstration data.
- Repository: github.com/Gunasekaranravieng/real-time-iot-streaming-lakehouse

Financial Fraud & Risk Lakehouse

Tech: PySpark, Apache Spark, Delta Lake, Medallion Architecture

- Built a synthetic financial-transaction Lakehouse with rule-based fraud detection, customer risk scoring and incremental transaction processing.
- Implemented Gold fraud/risk metrics, data-quality validation, reconciliation, audit logging and monitoring.
- Repository: github.com/Gunasekaranravieng/financial-fraud-risk-lakehouse

Enterprise Retail Lakehouse

Tech: PySpark, Spark SQL, Delta Lake, Unity Catalog

- Implemented Bronze/Silver/Gold retail processing with incremental Delta merges, reusable transformations and parameterised pipeline patterns.
- Added reusable data-quality checks, audit trails and error-handling workflows.
- Repository: github.com/Gunasekaranravieng/Enterprise_Retail_Lakehouse_Pipeline

EDUCATION

Bachelor of Science | Kamaraj University, Tamil Nadu | 2017 - 2020