

Luke Acosta

Junior Data Engineer

☎ (312) 555-0186 ✉ luke.acosta.data@example.com 💻 linkedin.com/example/lukeacosta 📍 Chicago, IL 60611

SUMMARY

Data engineer with five years of experience building dependable pipelines, warehouse capabilities, and production data services for research, analytics, and financial teams. Hands-on work covers Python, **Unix/Linux**, SQL, Postgres, MongoDB, AWS, Git, GitHub, Apache Airflow, Hadoop, Java, CI/CD, and normalized or denormalized data models. Translates requirements from portfolio managers, **quantitative developers**, analysts, vendors, and business partners into documented datasets and reliable workflows. Improves data quality through cleansing, anomaly detection, monitoring, alerting, lineage, **query optimization**, **performance tuning**, and clear operational runbooks. Supports production releases with careful testing, responsive troubleshooting, and practical communication. Ready to contribute technical judgment, curiosity, and steady collaboration to a data team supporting systematic trading and meaningful business decisions.

EXPERIENCE

Data Engineer

Cedar Peak Analytics 📅 January 2025 - Present 📍 Chicago, IL

Owns ingestion, warehouse improvements, validation services, and production workflows for market, reference, and portfolio datasets. Partners with portfolio analytics and quantitative development teams across AWS and Postgres environments, guiding delivery criteria, release quality, and operational support.

- Expanded **AWS** and Postgres warehouse sources for quantitative research datasets and portfolio **analytics**.
- Translated partner requirements into normalized staging models, mappings, and documented delivery criteria.
- Reworked Airflow dependencies and incremental loads for resilient vendor-file processing during source outages.
- Built **Python** checks for duplicates, stale prices, schema drift, and null patterns with monitored alerts.
- Tuned **Postgres** plans, indexes, and partitions while preserving denormalized analyst-facing views.
- Added **GitHub** pull-request checks and **CI/CD** gates for traceable, lower-risk pipeline releases

Data Engineering Associate

Blue Harbor Markets 📅 July 2023 - December 2024 📍 New York, NY

Developed **data pipelines** and warehouse models for quantitative research, combining vendor feeds, security master records, and transaction data. Supported Airflow operations, MongoDB checks, Postgres reconciliation, Hadoop profiling, AWS migration planning, vendor coordination, and restricted-data entitlement records.

- Consolidated vendor feeds with Python and **SQL** for **quantitative developers** conducting historical analysis.
- Modeled security master and transaction records with **normalization** rules and denormalized access layers.
- Added **MongoDB** checks and Postgres reconciliation queries for identifiers, duplicates, and timing differences.
- Implemented Airflow retries and alert routing for overnight workflows with clearer delivery visibility.
- Profiled **Hadoop** extracts and documented AWS storage and compute migration considerations for engineering leaders.
- Resolved vendor delivery anomalies and maintained entitlement records for restricted financial datasets

Data Engineering Analyst

Redwood Financial Research 📅 August 2021 - June 2023 📍 Boston, MA

Supported a small data platform team through supervised pipeline development, financial-**data cleansing**, warehouse reconciliation, Java utilities, Airflow experiments, anomaly checks, and Git-based delivery. Built practical **production support** habits while helping senior engineers trace records and improve recurring research workflows.

- Cleansed financial time series and reference files with **Python** and **Unix/Linux** before warehouse loading.
- Wrote Postgres transformations and reconciliation reports that traced source records through curated datasets.
- Contributed **Java** utilities for schema inspection and batch-file handling within a data platform.
- Created Airflow workflows with documented dependencies, ownership, retries, and validation steps.
- Applied anomaly thresholds and completeness checks before downstream research analysis and reporting.
- Used **Git** branching, code review, and unit tests for maintainable pipeline updates

PROJECTS

Cross-Market Data Quality Service 📅 2024

Built a Python service that checked schema drift, stale prices, duplicate observations, and unexpected nulls across financial feeds. Routed exceptions through monitoring queues and documented response steps for analysts and engineers.

Warehouse Performance Modernization 📅 2023

Improved Postgres access for recurring research workloads through query-plan review, indexing, partition strategies, and denormalized views. Shared findings with quantitative developers and documented repeatable tuning practices.

LEADERSHIP & AWARDS

- Received Cedar Peak Analytics Engineering Recognition for dependable warehouse releases and responsive production support.
- Selected by peers at Blue Harbor Markets as a trusted resource for vendor-data troubleshooting and entitlement records.
- Earned team recognition at Redwood Financial Research for clear pipeline documentation and thoughtful code reviews.

EDUCATION

Bachelor of Science in Computer Science

University of Massachusetts Amherst 🎓 GPA: 3.7 📅 2021 📍 Amherst, MA

Coursework: Data Structures, Database Systems, Operating Systems, Distributed Systems, Algorithms

CERTIFICATIONS

- AWS Certified Cloud Practitioner 📅 2024
- Apache Airflow Fundamentals, Astronomer Academy 📅 2023

TECHNICAL SKILLS

- **Programming Languages:** Python, Java, SQL
- **Operating Systems:** Unix/Linux, Ubuntu, Bash
- **Relational Databases:** Postgres, SQL, Query optimization
- **NoSQL Databases:** MongoDB, Document checks, Indexing
- **Cloud Platforms:** AWS, S3, EC2
- **Workflow Orchestration:** Apache Airflow, DAGs, Retry policies
- **Version Control:** Git, GitHub, Branching
- **CI/CD Tools:** GitHub Actions, Automated tests, Pull requests
- **Data Modeling:** Normalization, Denormalization, Schema design
- **Data Quality:** Data cleansing, Anomaly detection, Completeness checks
- **Monitoring Systems:** Alerting, Operational queues, Runbooks
- **Distributed Data:** Hadoop, Distributed extracts, Batch processing

SKILLS

- | | | | |
|--------------|------------|------------------|----------------------|
| • Python | • Postgres | • Apache Airflow | • Data cleansing |
| • SQL | • MongoDB | • Git | • Anomaly detection |
| • Java | • AWS | • GitHub | • Query optimization |
| • Unix/Linux | • Hadoop | • CI/CD | |

PROFESSIONAL AFFILIATIONS

- Member, Association for Computing Machinery, with interest in data engineering and dependable software delivery.
- Participant, Data Management Association community, sharing practical ideas on lineage, quality, and warehouse design.

LANGUAGES

- English (Native)
- Spanish (Intermediate)

ADDITIONAL INFORMATION

Work Status : Authorized to work in United States. No sponsorship required.

REFERENCES

AVAILABLE ON REQUEST