

# YASHAR MANSOURI

## Data Engineer

---

- Portfolio: [yasharmansouri.github.io](https://yasharmansouri.github.io)
- Phone: (202) 602-8464
- [yasharmansouri99@gmail.com](mailto:yasharmansouri99@gmail.com)
- [linkedin.com/in/yasharmansouri](https://linkedin.com/in/yasharmansouri)

## PROFESSIONAL SUMMARY

---

Results-driven Data Engineer with 5+ years of experience architecting enterprise data solutions and leading technical initiatives across consulting and engineering roles. Delivered scalable data platforms for Fortune 500 clients including Meta, Apple, Mayo Clinic, and government organizations, driving measurable improvements in data quality, system performance, and operational efficiency. Expertise in designing cloud-native architectures (AWS, GCP), implementing automated ETL workflows, and rapidly adopting emerging technologies to solve complex business challenges. Proven track record leading cross-functional teams, mentoring engineers, and translating technical innovation into business value across healthcare, technology, finance, and public sectors.

**Core Competencies:** Data Engineering • Solution Architecture • Cloud Platforms (AWS, GCP) • ETL/ELT • Python • SQL • Airflow • PostgreSQL • dbt • Kafka • Infrastructure as Code • Docker • Business Intelligence (Tableau, Looker) • DevOps & CI/CD • Technical Leadership • Client Delivery • Stakeholder Management • Data Governance

## PROFESSIONAL EXPERIENCE

---

### Data Engineer | DC Public Charter School Board (Contractor)

Washington, DC | February 2022 – September 2025

- Architected and maintained 75+ active Apache Airflow DAGs processing 10M+ daily records across student attendance, outcomes, assessments, enrollment projections, performance metric calculations and compliance reporting with 99%+ uptime serving 68+ charter schools
- Administered RDS PostgreSQL databases via role-based access control, automated per schema/object permission management; optimized junior developer SQL queries and diagnosed performance bottlenecks resolving concurrent task execution and improving database throughput by 50-80%
- Resolved 100+ emergency production issues including DAG and syncing failures and 77+ technical debt items through code refactoring, modernizing, error handling, diagnostics, and performance tuning, reducing incident response time by 60%
- Developed automated file processing infrastructure handling for 50+ different types of data including student records and financial reporting (\$500M+ annual budgets) through automated error handling, logging, validation frameworks and quality constraints
- Managed enterprise BI and low-code platforms (Looker, QuickBase, Tableau Cloud) serving 200+ school/staff users with role-based permissions and access controls; migrated 12+ Looker financial dashboards to Tableau reducing licensing costs by 70%
- Mentored 4+ analytic engineers through issue scoping, code reviews, 1:1s, technical guidance; generated architecture documentation, standard operating procedures and code review policy for organization handbook
- Served as sole data architect for organization, designing and implementing 70+ production data models including ASPIRE (Annual School Performance Index Report & Evaluation), School Reform Act (SRA) compliance reporting, Truancy, and financial data using SQLAlchemy ORM and Alembic migrations; maintained 3rd Normal Form standards across all schemas while developing FERPA-compliant dimensional data mart for staff analytics

## AI & Data Engineering Consultant | Deloitte

### McLean, VA | December 2021 – April 2023

- Delivered data engineering solutions across 4 Fortune 500 and enterprise client engagements (Apple, Meta, Mayo Clinic, Deloitte Internal) spanning ML infrastructure, privacy analytics, cloud architecture, and simulation technologies
- **Mayo Clinic:** Designed and implemented machine learning pipeline POCs on GCP for client data science teams utilizing GCP Vertex AI; created technical documentation and training materials for data science teams on ML pipeline deployment and best practices
- **Meta:** Developed and maintained privacy incident monitoring dashboards in Unidash and automated biweekly scorecard reports using Python and Presto SQL, reducing manual burden and incident response time; onboarded 3 junior data engineers to client's analytics team through shadowing sessions and comprehensive knowledge transfer documentation
- **Apple:** Built automated monthly/quarterly reporting pipelines for client's Wallet, Payments & Commerce team using Teradata and Trino SQL, enabling on-time delivery of monthly and quarterly financial KPIs to executive leadership; conducted User Acceptance Testing (UAT) during Teradata-to-Trino migration, identifying and resolving 10+ critical data quality issues preventing accurate revenue reporting
- **Deloitte:** Engineered Python-based synthetic data generation extension for NVIDIA Omniverse Isaac Sim POC, enabling randomized 3D robotics scenarios for machine learning model training in Deloitte's Enterprise Solutions practice

## Data Engineering Specialist | DC Public Charter School Board

### Washington, DC | May 2021 – December 2021

- Engineered 10+ automated data pipelines via Airflow DAGs including quickbase-to-database batch synchronization, staging-to-warehouse ELT, and infrastructure maintenance; managed AWS infrastructure (RDS, EC2, S3) with 99.9% uptime
- Designed normalized SQLAlchemy-to-PostgreSQL warehouse schemas and tables for 10+ different data sources supporting schools and internal staff business requirements; Developed QuickBase and Looker dashboards, reports, and forms enabling self-service analytics for schools, data analysts and finance team
- Led audit and code quality initiatives across different GitLab repositories, consolidating 16+ Looker repos and necessary libraries inside the main codebase, deprecating legacy modules, and updating documentation, reducing code duplication, merge request reviews and maintenance overhead
- Established unified data integrity and logging frameworks across 15+ ingestion pipelines automating data quality checks, data type conversion, error handling, and user notifications; developed comprehensive staff-facing data quality monitoring dashboards on user submitted files
- Administered Looker instance via managing role based permissions, content validation and query performance tuning

## Data Engineering Intern | DC Public Charter School Board

### Washington, DC | March 2021 – May 2021

- Contributed to PostgreSQL schema design across state assessments, student attendance, and equity metrics; Automated validation pipelines for 20+ DAGs
- Improved data model component generation using SQLAlchemy and Jinja2 templating reducing manual modeling overhead
- Supported database provisioning, troubleshooting pipeline failures, and legacy system migrations

# Data Science Intern | Allwyn Corporation

Herndon, VA | January 2020 – October 2020

- Developed machine learning models predicting hospital readmission rates with 82% accuracy using Gradient Boosting and Deep Learning; engineered ETL scripts extracting insights from 50M+ medical records (CDC, NIH, HCUP)
- Optimized data pipelines with PySpark reducing processing time by 60%; built Alteryx workflows for automated data transformation and statistical analysis
- Authored 3 technical white papers on machine learning methodologies and healthcare analytics best practices; mentored junior data scientists with project workflows

## TECHNICAL SKILLS

---

- **Programming / Scripting:** SQL, Python, Bash, YAML
- **Data Engineering & Orchestration:** Airflow, dbt, Spark, Kafka, ETL/ELT, Data Modeling, Stream Processing, Data Quality Frameworks, Data Governance
- **Databases & Data Warehouses:** PostgreSQL, Snowflake, Databricks
- **Business Intelligence / Data Visualization:** Tableau, Looker (LookML), QuickBase
- **Infrastructure & DevOps:** Docker, Kubernetes, Terraform, Infrastructure as Code (IaC), Git, GitHub, GitLab, pytest, Jenkins
- **AWS:** S3, EC2, RDS, Lambda, Glue, Athena, Step Functions, EMR, Kinesis
- **GCP:** Cloud Storage, BigQuery, Compute Engine, Cloud SQL, Dataflow, Cloud Composer, Vertex AI
- **ML/AI Frameworks:** TensorFlow (TFX), Scikit-learn, PyTorch
- **Languages:** English (Fluent), Persian (Native), Turkish (Fluent)

## EDUCATION

---

- **Master of Science in Data Science** | American University, Washington, DC | December 2021
- **Data Science Bootcamp Fellow** | Flatiron School, Washington, DC | December 2019
- **Master of Science in Biotechnology** | American University, Washington, DC | December 2017
- **Bachelor of Science in Molecular and Cellular Biology, Genetics** | Azad University, Tehran, Iran | February 2012

## CERTIFICATIONS

---

- Please check certifications posted on [linkedin.com/in/yasharmansouri](https://www.linkedin.com/in/yasharmansouri)