Yashar Mansouri

Senior Data Engineer

- Website: *yasharmansouri.github.io*
- Email: *yasharmansouri99@gmail.com*
- Phone: *(202) 602-8464*
- LinkedIn: *linkedin.com/in/yasharmansouri*
- Certifications: *linkedin.com/in/yasharmansouri/details/certifications*
- Book a meeting with me: *calendly.com/yasharmansouri/30min*

SUMMARY

Senior Data Engineer with 6+ years of experience building scalable data products that collect, process, and govern high-volume data for analytics, AI, and BI. Skilled in AI/ML workflows, data warehousing, dimensional and ER data modeling, database administration, and performance optimization using distributed processing frameworks. Proficient in Python and SQL with hands-on experience in batch and streaming pipelines, including CDC, incremental loads, and event-driven architectures. Proven ability to drive cross-functional projects autonomously, mentor engineers, and deliver reliable solutions across healthcare, technology, finance, and public sectors.

PROFESSIONAL EXPERIENCE

Senior Data Engineer | DC Public Charter School Board

Washington, DC | February 2022 – September 2025

- Acted as sole data architect; designed 70+ production data models including ASPIRE,for school performance evaluation, compliance reporting, and financial analytics using ER and dimensional modeling best practices
- Architected and maintained 75+ Apache Airflow DAGs that processed 10M+ daily records across attendance, assessments, enrollment, compliance, financial reporting with 99%+ uptime
- Built REST API integrations and automated file processing infrastructure, handling 50+ data types with schema validation and data quality constraints
- Administered PostgreSQL databases with role-based access control; optimized queries and eliminated performance bottlenecks, improving throughput by 50-80%
- Reduced incident response time by 60% by refactoring code, improving error handling, and resolving technical debt
- Managed enterprise BI platforms (Looker, QuickBase, Tableau Cloud) that served 200+ users across 130+ charter schools that facilitated critical funding decisions and policy compliance
- Migrated 12+ financial dashboards from Looker to Tableau, reducing licensing costs by 70%
- Partnered with analysts and stakeholders to gather requirements and translate business needs into technical specifications
- Mentored 4+ junior engineers through code reviews; authored architecture documentation and SOPs for the organizational handbook

AI & Data Engineering Consultant | Deloitte

McLean, VA | December 2021 – April 2023

- **Mayo Clinic**: Designed batch and streaming ML pipelines on GCP using Dataflow, Pub/Sub, BigQuery, and Vertex AI for data science teams; created technical documentation and training materials on deployment best practices
- **Meta**: Built privacy incident monitoring dashboards and automated biweekly scorecard reports using Python and Presto SQL; implemented data governance workflows for sensitive data handling; onboarded junior engineers through shadowing and knowledge transfer documentation
- **Apple**: Automated financial reporting pipelines for Wallet, Payments & Commerce team using Teradata and Trino SQL; collaborated with finance stakeholders to resolve 10+ critical data quality issues during Teradata-to-Trino migration
- **Deloitte**: Engineered Python-based synthetic data generation extensions for NVIDIA Omniverse Isaac Sim, enabling randomized 3D robotics scenarios for ML model training; implemented CI/CD workflows for simulation pipeline deployment

Data Engineer | DC Public Charter School Board

Washington, DC | March 2021 – December 2021

- Developed Airflow DAGs for batch synchronization, ELT workflows, and infrastructure maintenance
- Managed AWS infrastructure (RDS, EC2, S3) with 99.9% uptime
- Created transactional and analytical PostgreSQL data models using SQLAlchemy and Alembic
- Built QuickBase and Looker dashboards, enabling self-service data entry and analytics for schools and the internal finance team

- Led audit initiatives, consolidating 15+ Looker repos into the main codebase, reducing maintenance overhead
- Established unified data validation and logging frameworks, automating quality checks and error notifications

Data Scientist (Intern) | Allwyn Corporation

Herndon, VA | January 2020 – October 2020

- Developed AI/ML models predicting hospital readmission rates with 82% accuracy using scikit-learn and XGBoost
- Engineered ETL pipelines processing 50M+ medical records and generating predictive features using PySpark on Databricks
- Optimized data processing workflows, reducing pipeline runtime by 60%
- Authored 3 technical white papers on healthcare analytics and mentored 10+ graduate capstone students on AI workflows

Data Analytics Teaching Assistant | The George Washington University

Washington, DC | January 2020 – April 2020

- Instructed 20+ students on data analysis techniques and visualization best practices on Python, R, SQL, and Tableau
- Provided technical guidance on capstone projects, including data pipeline design and statistical analysis
- Developed grading rubrics and feedback frameworks to standardize assessment across assignments

Data Analyst | A2 Web Studios

Centreville, VA | May 2017 – March 2019

- Built and maintained dashboards and KPI reports using SQL and Excel; led migration of manual Excel reporting to Power BI, enabling automated refreshes and self-service analytics
- Designed data management and governance processes for the Data Engineering team
- Created automated data pipelines integrating internal and external data sources using Python and Jupyter Notebooks
- Collaborated with the Data Science team to optimize feature engineering and automation workflows
- Presented complex analyses to technical and non-technical stakeholders

TECHNICAL SKILLS

- **Languages & Scripting:** *Python, SQL, Bash*
- **Data Engineering & Orchestration:** *Airflow, Dagster, dbt, Spark/PySpark, Kafka, Flink, ETL/ELT Pipelines, CDC, Airbyte, Fivetran, REST APIs, FastAPI*
- **Data Platforms & Query Engines:** *Snowflake, Databricks, BigQuery, Redshift, Microsoft Fabric, Trino, Presto, Athena, Spark SQL*
- **Databases & NoSQL:** *PostgreSQL, Teradata, MongoDB, Redis, Cassandra, DynamoDB, Elasticsearch*
- **Data Modeling:** *Dimensional, 3NF, ER, Star/Snowflake Schema, Data Vault, Medallion Architecture*
- **Lakehouse & Data Formats:** *Delta Lake, Apache Iceberg, Hudi, LakeFS, Parquet, Avro, Polars, Apache Arrow*
- **Data Quality & Governance:** *Great Expectations, Soda, Data Contracts, Data Lineage, OpenLineage, DataHub, OpenMetadata*
- **AWS:** *S3, EC2, RDS, Lambda, Glue, Athena, Step Functions, EMR, Kinesis, Redshift, DynamoDB*
- **GCP:** *Cloud Storage, BigQuery, Compute Engine, Cloud SQL, Dataflow, Pub/Sub, Cloud Composer, Vertex AI*
- **Infrastructure & DevOps:** *Docker, Kubernetes, Terraform, Git, GitHub Actions, GitLab CI/CD, Infrastructure as Code (IaC)*
- **BI & Analytics:** *Tableau, Looker, Power BI, QuickBase*
- **ML & AI:** *Vertex AI, MLflow, Scikit-learn, XGBoost, Feature Engineering, Large Language Models (LLMs), Hugging Face, OpenAI API, LangChain, Retrieval Augmented Generation (RAG), Vector Databases, Prompt Engineering*
- **Spoken Languages:** *English (Fluent), Persian (Native), Turkish (Conversational)*

EDUCATION

- **Master of Science in Data Science** | American University | Washington, DC | January 2019 - December 2021
- **Data Science Bootcamp Fellow** | Flatiron School | Washington, DC | August 2019 - December 2019
- **Master of Science in Biotechnology** | American University | Washington, DC | August 2015 - December 2017
- **Bachelor of Science in Molecular and Cellular Biology, Genetics** | Azad University Tehran Medical Sciences | Tehran, Iran | September 2007 - February 2012