BITA ASHOORI
DATA ENGINEER | CLOUD DATA SPECIALIST (AZURE DATABRICKS / ETL)

📍Vancouver, BC 🔗 linkedin 🔗 portfolio website 🌐 github ✉ bitaashoori20@gmail.com 📞 (604) 724-6960

PROFESSIONAL SUMMARY

Data Engineer with 5+ years of experience designing and optimizing ETL/ELT pipelines, data warehouses, and analytics solutions across healthcare, retail, and e-commerce. Skilled in SQL, Python, SSIS, Airflow, and AWS, with strong expertise in data modeling (Star Schema, SCD2, CDC), data quality, and governance-driven development. Currently expanding knowledge in Azure Databricks, Delta Lakehouse, and Azure Data Factory to align with modern cloud data platforms. Passionate about building scalable, secure, and reliable data ecosystems that empower analytics and business insights.

TECHNICAL SKILLS

- Cloud & Data Platforms: Azure Databricks (Delta Lake, Unity Catalog in progress), Azure Data Factory, AWS (S3, Redshift, Lambda), SQL Server,
- Programming & Tools: Python (PySpark, pandas), SQL, T-SQL, SSIS, SSRS, Airflow, Git/GitHub Actions, Docker,
- Data Engineering: ETL/ELT design, SCD2/CDC, Star modeling, API integration, Data Quality, Logging & Validation
- DevOps & Governance: Azure DevOps CI/CD, Terraform (basic), Security & Access Control, Data Lineage, Monitoring & Alerting
- Visualization & BI: Power BI, SSRS, Denodo (Virtualization)

PROFESSIONAL EXPERIENCE

Data Engineer at O2E Brands, Vancouver, BC (Remote)   2021 – 2024
- Designed and deployed end-to-end ETL pipelines integrating diverse sources into AWS Redshift for analytics.
- Developed Python/Airflow workflows for batch and near-real-time ingestion with incremental MERGE and SCD Type 2 logic.
- Modernized legacy ETL into modular Airflow DAGs, improving reliability, observability, and maintainability across environments.
- Refactored complex SQL stored procedures, boosting performance by 60%.
- Optimized SQL and Python transformations, reducing execution time by 30%.
- Partnered with BI and data science teams to define data models, governance standards, and validation frameworks.
- Implemented monitoring, alerting, and logging, cutting pipeline failures by 70%.

Data /ETL Developer at Vancouver Coastal Health, Vancouver, BC (On-site)   2019 – 2021
- Designed and maintained data warehouse solutions using SQL Server and SSIS, implementing SCD Type 2 and audit frameworks for regulatory compliance.
- Built ETL pipelines integrating patient, clinical, and operational data from multiple systems into a governed centralized repository.
- Developed optimized T-SQL procedures, views, and transformations, improving report performance and reliability.
- Collaborated with clinical, operations, and IT teams to enhance data governance, lineage tracking, and quality control processes.

Database Administrator, Georgia Main Food Group, Burnaby (On-site)   2017 – 2019
- Administered and optimized SQL Server databases supporting POS, finance, and inventory systems.
- Managed backups, restores, indexing, and security roles, ensuring data integrity and high availability.
- Built and maintained SSIS ETL packages and SSRS / Power BI dashboards for enterprise reporting and performance monitoring.
- Collaborated in migration of legacy databases to SQL Server 2019, improving performance, compliance, and maintainability.

Data Pipeline with Databricks : Azure Databricks + PySpark + Python + SQL
- Designed and implemented a modern ETL pipeline following the Medallion Architecture (Bronze, Silver, Gold) to transform raw data into analytics ready assets.
- Built and optimized incremental data pipelines using PySpark and Spark SQL, ensuring data quality and reliability.
- Developed Silver and Gold layer Delta tables for data cleansing, enrichment, and aggregation.
- Applied data governance principles using Unity Catalog and built visual dashboards in Databricks SQL / Power BI Service.
- Tools: Azure Databricks, Delta Lake, PySpark, Spark SQL, Unity Catalog, Power BI Cloud

Real-Time Marketing Pipeline : Databricks + PySpark + Delta Lake
- Built a real-time streaming pipeline to process marketing events using PySpark Structured Streaming and Delta Lake.
- Implemented Bronze,Silver,Gold architecture with schema enforcement and checkpointing for reliability.
- Automated deployment and data validation via Airflow DAGs and GitHub Actions CI/CD.
- Reduced reporting latency from hours to minutes and improved data consistency across dashboards.
- Tools: Databricks, Airflow, GitHub CI/CD

PySpark Sales Pipeline – Delta Lake Optimization
- Developed a production style PySpark ETL pipeline to ingest and transform large sales datasets into Delta Lake.
- Implemented partitioning and optimization strategies to enhance performance and query efficiency.
- Achieved ~40% faster transformations and improved reporting accuracy using Delta Lake optimizations.
- Deployed and tested on Databricks (local) with AWS S3 storage integration.
- Tools : PySpark, Delta Lake, AWS S3

- **Portfolio**: https://bashoori.github.io/data-engineering-portfolio/

EDUCATION

- B.A.Sc. in Computer Science & Mathematics, Azad University
- Applied Database Administration and Design, BCIT (2015–2017)
- Databricks Academy Accreditations (2025): Fundamentals • Get Started for Data Engineers • Generative AI Fundamentals
- PySpark Essential Training: Building Data Pipelines - LinkedIn Learning (2025)
- End-to-End Data Pipeline with Databricks - LinkedIn Learning (2025)
- Mastering Azure Databricks for Data Engineers Specialization - Coursera (in progress)