

Abhay Dawar

British Columbia, Canada | 📞 +1-778-281-6969 | ✉ iamabhaydawar@gmail.com | [in linkedin.com/in/iamabhaydawar](https://www.linkedin.com/in/iamabhaydawar) | github.com/iamabhaydawar

PROFESSIONAL SUMMARY

Data Engineer with hands-on experience designing, building, and maintaining secure data processing pipelines using Microsoft Azure data services and Databricks. Proven ability to build databases, data marts, and data warehouses with ETL operations, data validation, and data quality enforcement. Strong background in data governance, privacy, and security compliance within healthcare and regulated environments. Collaborative team player experienced in working with data analysts, business analysts, and stakeholders to deliver reliable, well-documented data solutions that drive reporting and actionable business insights.

EDUCATION

Indian Institute of Technology, Madras

Chennai, India

Bachelor of Science in Data Science & AI — CGPA: 8.8/10

College of New Caledonia

British Columbia, Canada

Diploma in Business Administration - Economics & Finance — CGPA: 3.1/4.33

TECHNICAL PROJECTS

Healthcare DLT Medallion Architecture Pipeline — Azure Databricks, Delta Live Tables, PySpark, ADLS, SQL

- Designed and built a secure data processing pipeline using Databricks Delta Live Tables with Bronze-Silver-Gold medallion architecture, processing 500GB+ healthcare patient records with role-based access controls
- Implemented data validation and cleansing procedures with quality constraints ensuring integrity and dependability of clinical data; built automated quality metrics monitoring and schema evolution tracking
- Ensured compliance with healthcare data governance, privacy, and security standards; developed algorithms to transform raw data into useful data models and built multi-dimensional analytics tables for hospital operations and compliance reporting

🔗 github.com/iamabhaydawar/HealthCare_DLT_Medallion_Pipeline

Travel Booking SCD2 Data Warehouse — Azure Synapse, Azure Data Factory, Cosmos DB, PySpark, SQL

- Built a data warehouse with star schema dimensional modeling and SCD Type 2 for historical customer tracking; designed source-to-target mapping with surrogate keys and temporal columns for full audit trail
- Created ETL (Extract, Transform, Load) operations with PySpark for data ingestion, PyDeequ quality validation, and incremental processing; implemented idempotent Delta MERGE operations for reliable data delivery
- Improved scalability, efficiency, and cost-effectiveness with Z-ORDER clustering and statistics analysis; built reporting and analytical tools providing actionable insight into key business performance metrics

🔗 github.com/iamabhaydawar/Travel_Booking_SCD2_Warehouse_Project

E-Commerce Event-Driven Databricks Pipeline — Databricks, Spark, PySpark, Delta Lake, Azure Event Hub

- Designed and implemented data pipelines for data ingestion, processing, and transformation with incremental loading, Delta Lake ACID transactions, and automated file archival for data consistency
- Built modular ETL stages for orders, customers, products, inventory, and shipping with comprehensive data validation, business rule checks, and error handling across all staging tables
- Monitored and resolved data pipeline problems to ensure consistency and availability; implemented Delta MERGE upsert operations with data lineage tracking and idempotent processing for business analytics

🔗 github.com/iamabhaydawar/Ecomm_event_driven_dbx_Pipeline

PROFESSIONAL EXPERIENCE

Financial Data Analyst

Jan. 2025 – Dec. 2025

DNW Distribution

British Columbia, Canada

- Collaborated with data analysts, business analysts, and business SMEs to understand data needs; built ETL pipelines consolidating financial, sales, and logistics data from internal systems and Costco platforms
- Performed root cause analysis on internal and external data to answer specific business questions and identify opportunities for improvement; engineered data models with SQL/PySpark for forecasting and operational reporting
- Identified and implemented internal process improvements including optimizing data delivery and automating manual processes; planned, prioritized, and managed workload within a time-sensitive environment
- Created BI dashboards and reporting tools providing actionable insight into key business performance metrics; applied ML improving decision accuracy by 40%; presented results using audience-appropriate communication

TECHNICAL SKILLS

Cloud & Data Platforms: Microsoft Azure (Fabric, Synapse, Data Factory, Azure SQL Database, Azure Data Lake, Azure Blob Storage, Event Hub, Cosmos DB), Databricks, Delta Lake, Snowflake, AWS (S3, Glue, Redshift, Lambda)

Data Engineering & ETL: ETL/ELT Design and Implementation, Data Pipelines, Data Ingestion/Processing/Transformation, Data Migration, SCD Type 1/2, CDC, Stream/Batch Processing, Delta Live Tables, Kafka, Airflow

Programming: Python, SQL (PostgreSQL, MySQL, SQL Server), PySpark, Java, JavaScript, TypeScript, Bash, C, C++, Go

Data Governance & Quality: Data Lineage, Data Dictionaries, Metadata Catalogs, Quality Frameworks, Privacy and Security Compliance, Governance Standards, Data Validation and Cleansing

Data Modeling: Star/Snowflake Schema, Medallion Architecture, ERD Design, Dimensional Modeling, Data Warehousing, Lakehouse Architecture, Data Product Design, Metadata Management

Databases: PostgreSQL, MySQL, SQL Server, MongoDB, Redis, Cassandra, Snowflake, Cosmos DB

DevOps: Docker, Kubernetes, Terraform, Git/GitHub, CI/CD (Azure DevOps, GitHub Actions), Linux/Unix

BI & Reporting: Power BI, Tableau, Looker, Excel (Advanced), DAX, Data Visualization