Apoorv Pratap Dhaygude

617-505-0704 | dhaygude.a@northeastern.edu | linkedin.com/in/apoorv-dhaygude/ | Boston, MA

Education

Northeastern University — Boston, MA
Master of Science in Information Systems — *Aug 2025*
Graduate Teaching Assistant: Data Science Engineering Methods and Tools

Savitribai Phule Pune University — Pune, India
Bachelor of Engineering in Computer Science Engineering — *Jul 2021*

Professional Experience

Data Engineer Co-op — July 2024 – Dec 2024
Johnson & Johnson MedTech — *Danvers, MA*

- Developed a C# CLI tool using .NET to convert EEPROM binary files to JSON with marshalling technique
- Addressed broader data retrieval needs by designing scalable solutions that process multiple values efficiently
- Standardized data formats across tools and platforms, ensuring consistency and reducing data discrepancies by 30%
- Implemented CRC checksum to detect and prevent file corruption, improving data accuracy to 99%
- Built a temporal RDBMS with JSON schemas, handling 1TB+ of medical study data, enabling 99.9% uptime
- Led ETL data mapping for GXP compliance, ensuring regulatory adherence and reducing audit errors by 15%
- Supported remediation efforts for 4+ platforms by ensuring quality and comprehensive documentation
- Tested internal APIs using Postman, resolving 50+ defects to improve accuracy by 20%
- Engineered a dataset for extreme test cases, increasing 15+ Tableau dashboards accuracy by 30%

Data Engineer — Aug 2021 – Jun 2023
Tata Consultancy Services — *Pune, India*

- Optimized ETL pipelines in Google BigQuery, enabling ingestion of 20M+ rows daily, improving performance by 25%
- Designed CI/CD ETL pipelines in GitLab for secure data transfer to the cloud, reducing errors & security risks by 20%
- Enhanced data warehouse performance by 45% through table partitioning, clustering, and materialized views
- Streamlined data flow pipelines, ensuring seamless connectivity between operational, analytics, and BI systems
- Executed scalable Apache Airflow DAGs, processing data across distributed systems with 99.9% reliability
- Re-engineered manual data processes, enabling scalability, and repeatability, reducing data processing time by 40%
- Devised Source-to-Target mapping (STTM) and metadata management, improving data lineage tracking
- Delivered business-critical insights through Looker Studio dashboards, resulting in a 25% boost in client acquisition

Technical Skills

Languages: Python, Java, C++, C#, JavaScript, Bash
Cloud Platforms: GCP (BigQuery, Dataflow), Microsoft Azure (Data Factory, Azure SQL), AWS (S3, Lambda)
Big Data & Distributed Systems: Apache Hadoop, Apache Spark, Apache Impala, Apache Airflow, Parquet
Databases: BigQuery, Relational Database, PostgreSQL, MariaDB, MSSQL, MongoDB, Neo4j
Certifications: Google Cloud - Associate Cloud Engineer, Microsoft - Azure Fundamentals

Academic Projects

Food Inspection Analysis | *Azure Data Factory, ETL/ELT* — Mar 2024 – Apr 2024

- Integrated data from 10+ disparate sources, improving data accuracy and consistency by 25%
- Created data workflows with Azure Data Factory, reducing integration time by 60%
- Synthesized ETL/ELT pipelines with 7+ calculated transformations for real-time insights
- Outlined Power BI semantic models to enable interactive analysis and dynamic inspection insights

Medical Device Sales Performance | *Snowflake, Apache Spark, Tableau* — Feb 2024 – Mar 2024

- Architected a Snowflake-based data warehouse to consolidate 100M+ sales, customer, and inventory records
- Sparked ETL pipelines using Apache Spark to ingest and transform 50GB+ of data, ensuring reliable processing
- Boosted query performance with clustering and materialized views, reducing query execution time by 30%
- Created 15+ interactive Tableau dashboards to analyze sales trends, increasing revenue insights accuracy by 30%

Hospital Management System | *PostgreSQL, Python, SQL* — Jan 2024 – Feb 2024

- Constructed a relational database to manage 500K+ patient records, staff details, and billing workflows for a hospital
- Improved query execution with partitioning and indexing, reducing response time for complex queries by 40%
- Generated 20+ advanced SQL queries to track resource utilization, patient treatments, and billing insights in real-time
- Produced role-based access controls, securing 500K+ patient records and ensuring regulatory compliance