Himanshu Singh

+91 9129491699 | dev88himanshu@gmail.com

[LinkedIn ↗](#) | [Portfolio ↗](#)

PROFESSIONAL SUMMARY

Azure Certified Data Engineer with **4+ years of experience** in building and optimizing **ETL/ELT pipelines**, **data warehouse modernization**, and **cloud data lakehouse solutions** on the **Azure ecosystem (ADF, Databricks, Synapse, ADLS)**. Skilled in **data modeling, data governance, CI/CD with Azure DevOps**, and applying **performance optimization techniques** that deliver up to **60% efficiency improvements**. Proficient in **Python, SQL, Shell Scripting, Spark, Hive, Impala, and MapReduce**, with proven success in processing **multi-terabyte datasets** for **Fortune 100 Healthcare and Retail clients**. Experienced in enabling **scalable, secure, and production-ready analytics platforms**.

TECHNICAL SKILLS

Cloud & Data Engineering:	ADF, Databricks, Azure Synapse, ADLS/Blob Storage, Data Lakehouse
Pipelines & Warehousing:	ETL/ELT, Data Modelling, Data Warehousing OLAP/OLTP
Programming:	Python, SQL, Shell Scripting
Big Data:	Hadoop, Spark, PySpark, RDD, Spark Streaming, Hive, Impala, MapReduce, YARN
Orchestration & Workflow:	Tivoli Workload Scheduler (TWS), Sqoop, Apache Airflow, Azure DevOps
Developer Tools:	ServiceNow, Git, WinSCP, PuTTY
GenAI:	LangChain, LlamaIndex, HuggingFace
Familiar:	Django (Rest API), MySQL, PostgreSQL, ChromaDB

EXPERIENCE

TATA CONSULTANCY SERVICES   KAISER PERMANENTE (Mar 2023 – Present)
Data Engineer   *Hyderabad, India*

- Designed and optimized **enterprise-scale data pipelines** using **Azure Data Factory (ADF)** to ingest and orchestrate **20TB+ healthcare datasets**, integrating them into **ADLS** and **Databricks Delta Lake** for advanced analytics and reporting.
- Automated **ETL/ELT workflows** with **ADF**, **ADLS**, **Databricks**, and **PySpark**, improving processing efficiency by **60%** and enabling **real-time data availability**.
- Delivered actionable **business insights** to stakeholders by improving data accuracy by **100%** through enhanced validation and quality frameworks.
- Integrated **on-premise, third-party, and Hadoop-based data sources** (**Oracle, Taleo, JSON, Avro, CSV, Parquet**) into **ADLS** via ADF connectors, enabling seamless enterprise-wide analytics.
- Applied **Spark performance tuning techniques** (caching, partitioning, broadcast joins), reducing processing time by **20%** for **2M+ daily records** across **Hadoop and Azure environments**.

TATA CONSULTANCY SERVICES   WALGREENS (Oct 2021 – Feb 2023)
Data Engineer   *Noida, India*

- Collaborated with business teams to deliver **enterprise reporting solutions** by processing large-scale datasets across ETL systems.
- Developed and optimized **MapReduce programs** within the **Hadoop ecosystem** for large-scale text processing, reducing execution time by **90%** (2 hours → 15 minutes).
- Automated **data workflows** using **Python**, **HiveQL**, and **Shell Scripting**, improving operational efficiency by **50%** and minimizing manual intervention.
- Performed **data quality, consistency, and reconciliation checks** using **Hive**, **Impala**, and **Spark**, improving trustworthiness of analytical datasets by **30%**.
- Streamlined **ETL orchestration** by automating complex pipelines with **Python**, **HiveQL**, and **Shell**, reducing manual efforts by **40%** and improving throughput by **60%**.
- Worked extensively on **Hadoop Distributed File System (HDFS)** for data ingestion and batch processing, laying the foundation for migration to modern **Azure-based architectures**.

CERTIFICATIONS

Generative AI (Udemy) | 2025   [View Certificate](#)

Databricks Fundamentals | 2025   [View Certificate](#)

Azure Fundamentals (TCS Digital) | 2025   [View Certificate](#)

Python Programming (TCS Digital) | 2022   [View Certificate](#)

Microsoft Certified: Azure Data Fundamentals (DP-900) | Expected 2025

PROJECTS

LeetCode SQL to PySpark (Top 50 Questions)   [View on GitHub](#)

DataLemur SQL to PySpark   [View on GitHub](#)

DataLemur SQL to PySpark (RDD Implementation)   [View on GitHub](#)

Databricks Ecommerce analytics   [View on GitHub](#)

Databricks ETL   [View on GitHub](#)

EDUCATION

NIET   Greater Noida, UP
Bachelor of Technology in Computer Science & Engineering   *Aug 2017 – Sep 2021*

ABC Public School   Gorakhpur, UP
Senior Secondary School   *Apr 2015 – Mar 2017*