

OJASWI

Data Engineer · Databricks · PySpark · Delta Lake · Medallion Architecture
• +49 160-5512-108 • ojaswishere@gmail.com • linkedin.com/in/ojaswishere • github.com/DevOGhere

EDUCATION

M.Sc. Applied Data Science Georg-August-Universität Göttingen, Germany	April 2026 – Present
B.Tech. Computer Science & Engineering Guru Gobind Singh Indraprastha University (GGSIPU) <i>Minor: Artificial Intelligence & Machine Learning</i> • <i>Relevant Coursework: Data Warehousing & Mining · Data Modelling · Probability & Statistics</i>	Nov 2021 – May 2025 GPA: 9.62/10 (≈1.2 German scale, <i>Sehr Gut</i>)

PROFESSIONAL EXPERIENCE

Veersa Technologies <i>Data Engineer</i> <i>Data Engineering Intern</i>	Sep 2024 – April 2026 <i>Aug 2025 – April 2026</i> <i>Sep 2024 – Jul 2025</i>
Promoted within 11 months; independently owned 3–4 concurrent client pipelines end-to-end on Databricks/Delta Lake over 18 months, delivering production healthcare analytics to C-suite.	
• Architected an Agent Productivity Analytics Platform integrating 5+ heterogeneous sources (Pega, Siebel, Amazon Connect, UKG, SharePoint) into a PySpark + Delta Lake medallion pipeline on Databricks; owned stakeholder communication up to C-suite level. <i>Awarded Rising Star Q3 '25</i>. • Engineered a specialty pharmacy fulfillment pipeline for a major global pharmaceutical client – gold-layer fact table with multi-join SQL across Snowflake + Databricks, SCD/watermark-based ingestion, processing millions of records daily with ACID-compliant Delta Lake transactions. • Diagnosed and fixed a silent data loss incident where an upstream schema change had been dropping thousands of Bronze→Silver records undetected for months; engineered backward-compatible transformation logic and fully backfilled all affected historical data. • Built a configuration-driven ingestion framework cutting new client onboarding from 2 weeks to 3 days; automated SQL-based data quality checks with Databricks Alerts and orchestrated all pipelines via Databricks Jobs across production healthcare datasets.	
Indika AI, AI/ML Intern	Jul 2024 – Aug 2024
• Fine-tuned OpenAI Whisper via HuggingFace Transformers for Indian language ASR, achieving 30–40% reduction in Word Error Rate (WER) across low-resource languages. • Developed an automated audio extraction pipeline using YouTube Data API with precise timestamp control, generating high-quality training datasets for ASR model optimization.	

KEY COMPETENCIES

Programming	Python, SQL, PySpark, C/C++ (DSA)
Data Engineering	Databricks, Delta Lake, Apache Spark, AWS S3, Snowflake, Medallion Architecture, SCD Type 1 & 2, Data Lakehouse
CI/CD & Tooling	Azure DevOps, Git, Databricks Jobs
Monitoring & Quality	Databricks Alerts, SQL-based DQ Validation, Data Pipeline Monitoring
Viz & Libraries	Power BI, Pandas, NumPy, HuggingFace Transformers
Languages	English (C1) · German (A2) – <i>Werk in Progress</i> · Hindi (Native)

PUBLICATIONS

Kundra, R., & Ojaswi. “Assessing the Efficiency of Gradient Descent Variants in Training Neural Networks,” *Darpan International Research Analysis*, vol. 12, no. 3, pp. 596–604, Sep. 2024. DOI: 10.36676/dira.v12.i3.114