

MALIK ZOHAIB HASSAN

Junior Data Engineer - Azure Cloud & ETL Pipelines

zohaibmalikofficial@gmail.com | +491783576366 | Ilmenau, Deutschland | [linkedin.com/in/zohaibmalikofficial](https://www.linkedin.com/in/zohaibmalikofficial)

BERUFLICHES PROFIL

Highly motivated and detail-oriented Working Student with experience in project support, data analysis, and presentations. Proficient in **Python**, **SQL**, and MS Office Suite (especially **PowerPoint** and **Excel**), seeking to support the Chief of Staff to the CTO at Statista. Currently pursuing a Master's degree in Computer and Systems Engineering at TU Ilmenau, focusing on Cloud Computing and System Security. Eager to learn in a fast-paced environment and contribute to strategic initiatives within the Tech Office. Experience with collaboration tools like Jira and Confluence.

TECHNISCHE KOMPETENZEN

Data Engineering & ETL:

Azure Databricks, **Apache Spark (PySpark)**, ETL/ELT-Pipelines, Data Factory, Airflow, **Kimball Datenmodellierung**, Data Vault (Grundlagen)

Programmierung & Scripting:

Python (Fortgeschritten), **Scala (Grundlagen)**, SQL (Fortgeschritten), pandas, PySpark, pytest, Flask

Cloud & Azure Services:

Azure Data Lake Storage, **Azure SQL Database**, **Azure Synapse Analytics**, AWS (Übertragbar), Kubernetes, Docker

Datenbanken & Storage:

PostgreSQL, MySQL, **Azure SQL**, **Delta Lake**, Data Warehouse Design, Datenbankoptimierung, NoSQL (MongoDB)

Datenqualität & Monitoring:

Data Quality Frameworks, **Azure Monitor**, Datenvalidierung, Fehlerbehandlung, Performance-Tuning

Agile & Kollaboration:

Git, CI/CD, **Agile Methoden**, Stakeholder-Kommunikation, Technische Dokumentation, Deutsch (A2+→B1)

ZENTRALE DATA ENGINEERING PROJEKTE

Azure Databricks ETL-Pipeline für Retail-Analytics mit Kimball Dimensional Modeling

Azure Databricks, PySpark, Azure Data Lake Storage Gen2, Delta Lake, Azure SQL Database, Python

Entwicklung einer skalierbaren ETL-Pipeline zur täglichen Verarbeitung von Verkaufsdaten, Lagerbeständen und Kundendaten aus 8 verschiedenen Quellsystemen (APIs, CSV, JSON, Datenbanken). Implementierung von **Kimball Star-Schema** mit 3 Fact-Tables (Sales, Inventory, Returns) und 7 Dimension-Tables. **Optimierungstechniken:** Delta Lake für ACID-Transaktionen, Partitionierung nach Datum/Region, Z-Ordering für bessere Query-Performance. **Ergebnis:** 65% schnellere Abfragezeiten durch optimierte Tabellenstrukturen, 99.7% Datenqualität durch implementierte Validierungsregeln, Reduzierung der Pipeline-Laufzeit von 4h auf 45min durch Parallelisierung.

Data Lakehouse Architektur mit Bronze-Silver-Gold Medallion Pattern

Azure Data Factory, PySpark, Delta Lake, Azure Synapse Analytics, Apache Airflow, SQL

Design und Aufbau einer mehrstufigen Data-Lakehouse-Architektur nach Medallion-Pattern für E-Commerce-Daten: **Bronze** (Raw Data Ingestion), **Silver** (Cleansed & Validated), **Gold** (Business-Ready Analytics). Implementierung automatisierter ELT-Prozesse mit Azure Data Factory für 12 verschiedene Datenquellen. **Datenqualitätstechniken:** Schema-Validierung, Duplikaterkennung, referentielle Integrität, automatische Anomalie-Detection. **Ergebnis:** Zentrale Datenverfügbarkeit für 25+ Analytics-Teams, 80% Reduzierung der Datenbereitstellungszeit von 6h auf 1h, Self-Service Analytics für Business-User.

Multi-Source Data Integration mit Data Vault 2.0 Modellierung

Azure Data Factory, SQL Server, Python, Azure SQL Database, Data Vault 2.0, PowerBI

Integration heterogener Unternehmensdaten aus ERP-Systemen, CRM-Datenbanken und Excel-Dateien unter Verwendung von **Data Vault 2.0 Methodik**. Aufbau von Hubs (Geschäftsentitäten), Links (Beziehungen) und Satellites (Kontextdaten) für historische Datenaufbewahrung und Audit-Compliance. **ETL-Optimierungen:** Incremental Loading, Change Data Capture (CDC), parallelisierte Batch-Jobs. **Performance-Tuning:** Indexstrategien, Query-Optimierung, Partitionierung. **Ergebnis:** 100% Datenrückverfolgbarkeit für Compliance-Anforderungen, 70% Verbesserung der Reportinggenauigkeit, Reduzierung manueller Datenabgleiche von 16h/Woche auf 2h/Woche.

Automatisierte ETL-Pipeline mit Datenqualitäts-Monitoring

Apache Spark, Python, PostgreSQL, Apache Airflow, Great Expectations, Grafana

Entwicklung einer robusten ETL-Pipeline zur Verarbeitung von Transaktionsdaten aus Online-Payment-Systemen mit integriertem Datenqualitäts-Framework. Implementierung von **Great Expectations** für automatisierte Datenvalidierung, Schema-Evolution-Handling und statistische Anomalie-Detection. **Monitoring & Alerting:** Grafana-Dashboards für Pipeline-Performance, Slack-Integration für Fehler-Benachrichtigungen, automatische Wiederanlauf-Mechanismen.

Skalierungstechniken: Spark-Cluster-Optimierung, Memory-Management, broadcast joins für kleine Lookup-Tabellen.

Ergebnis: 99.5% Pipeline-Verfügbarkeit, 90% Reduzierung manueller Datenvalidierung, automatische Behandlung von 95% der auftretenden Datenqualitätsprobleme.

Kubernetes-basierte Microservices Data Processing Platform

Kubernetes, Docker, Python Flask, Apache Kafka, Redis, PostgreSQL, Prometheus

Co-Entwicklung einer containerisierten Datenverarbeitungsplattform mit Custom Pod Autoscaler (CPA) für dynamische Skalierung basierend auf p95-Latenz-Metriken aus Prometheus. Design von 6 Microservices für verschiedene ETL-Aufgaben: Data Ingestion, Validation, Transformation, Quality Checks, und Output Generation. **DevOps-Techniken:** GitOps mit ArgoCD, CI/CD-Pipelines, Infrastructure as Code (YAML), Service Mesh für Inter-Service-Kommunikation. **Monitoring & Observability:** Prometheus Metrics, Grafana Dashboards, distributed Tracing. **Ergebnis:** Horizontale Skalierung von 2-50 Pods je nach Last, 40% bessere Resource-Utilization im Vergleich zu CPU-basiertem HPA, 99.9% Service-Verfügbarkeit.

BERUFSERFAHRUNG

Data Engineer / WordPress Developer

Nov 2021 - Apr 2025

FixRunner

USA (Remote)

- Developed and optimized data processing workflows using PHP, Javascript and Python to enhance website functionality for clients.
- Collaborated remotely with global teams utilizing project management tools like Git, GitHub, and Asana to ensure efficient project delivery.
- Diagnosed and resolved website errors, plugin conflicts, and security vulnerabilities across numerous projects, contributing to improved website functionality and security.
- Documented technical implementations and communicated effectively with stakeholders to ensure clarity and transparency.
- Integrated third-party APIs and implemented automation workflows to streamline project processes and increase efficiency.

AUSBILDUNG

Master of Research in Computer and Systems Engineering

Technische Universität Ilmenau (TU Ilmenau)

Ilmenau, Deutschland | Okt 2024 - Laufend

Aktuelle Kurse: Cloud Computing, Software Safety, System Security , Advanced Algorithms, Mobile Communication Networks

Relevanter Fokus: Distributed Data Systems, Cloud-Architekturen, Sicherheit in der Datenverarbeitung , DevOps-Methoden, Enterprise Software

Bachelor of Science in Computer Science

COMSATS University Islamabad

Pakistan | Feb 2019 - Feb 2023

Kernfächer: Datenstrukturen, Algorithmen, Datenbankmanagement, Data Mining , Software Engineering

Abschlussprojekt: "Scalable Data Processing System" - Implementierung eines verteilten Datenverarbeitungssystems mit Apache Spark

ZERTIFIZIERUNGEN & WEITERBILDUNG

Azure & Data Engineering:

- Azure Data Fundamentals (DP-900) - In Vorbereitung
- Certified Data Engineer Associate

Technische Spezialisierungen:

- ETL/ELT Pipeline Development
- Advanced SQL für Data Warehousing
- Python for Data Engineering
- Cloud Data Architecture Patterns
- Agile Data Engineering Methoden

ZUSÄTZLICHE QUALIFIKATIONEN

Sprachkenntnisse & Kommunikation:

Deutsch: A2+ (Aktiv in B1-Kurs, Ziel: B2 bis Ende 2025)

Englisch: Fließend (Technisch & Geschäftlich)

Stakeholder-Kommunikation: Erfahrung mit interdisziplinären Teams und technischer Dokumentation

Verfügbarkeit & Bereitschaft:

Status: Autorisierte Arbeitserlaubnis in Deutschland

Startdatum: Oktober 2025 (flexibel)

Lernbereitschaft: Kontinuierliche Weiterbildung in Azure-Technologien

Teamarbeit: Agile Arbeitsweise, selbstorganisiert, proaktiv