

Elyas El Bastawissy

Data & AI Engineer · Databricks & Azure · Agentic AI & RAG

elyas@mithelfer.ai · 0152 21796145

Leibnizstraße 35a, 10625 Berlin · mithelfer.ai/profil



VERFÜGBARKEIT

Ab sofort · 4-5 Tage/Woche ·
Remote / bundesweit

SPRACHEN

Deutsch (C2) · Englisch (C1)

ERFAHRUNG

~6 Jahre Data Engineering
~3 Jahre Databricks

PROFIL

Data & AI Engineer mit sechs Jahren Data-Engineering-Erfahrung, davon drei Jahre produktive Databricks-Praxis, überwiegend auf Azure. Schwerpunkt: produktionsreife Lakehouse-Plattformen in regulierten Umgebungen (Banking, Versicherung) und agentische KI-Anwendungen im Enterprise-Umfeld (Handel).

Data Engineering: Ingestion-, Transformations- und Aufbereitungspipelines in PySpark, Delta Lake und SQL unter Unity-Catalog-Governance; konfigurations- und metadatengetriebene Onboarding-Frameworks, die heterogene Quellsysteme (REST- und SOAP-APIs, Dateifeeds, XML, PDF) über Konfiguration statt Code-Änderung anbinden; dimensionale Modellierung und SCD-Type-2-Historisierung entlang der Medallion-Architektur; Migration bestehender Data-Warehouse-Workloads nach Databricks mit dbt und Data Vault 2.0. Fundiertes Spark-Internals-Verständnis (Partitionierung, Shuffle-Reduktion, Skew-Behandlung, Join-Strategien) als Basis für Performance-Tuning und Fehlerdiagnose in produktiven Pipelines.

GenAI & Agentic AI: End-to-End-RAG-Pipelines (Parsing, Chunking, Embeddings, Vector Search, Re-Ranking) und Multi-Agent-Architekturen (Mosaic AI Agent Framework, Genie Spaces) mit MCP-Tool-Integration, bereitgestellt über einen zentral verwalteten, OpenAI-kompatiblen Endpoint und angebunden an ein selbst gehostetes Chat-Frontend. LLM-Bereitstellung, Prompt-Flow-Orchestrierung und Model-Monitoring über Azure AI Foundry und MLflow. Produktiver Einsatz KI-gestützter Entwicklungswerkzeuge (Claude Code, GitHub Copilot) für Entwicklung, Testing und Dokumentation.

Arbeitsweise: Infrastructure as Code (Terraform), gestufte CI/CD-Pipelines mit Plan-/Apply-Trennung und Approval-Gates, Git-basierte Entwicklung mit Code-Reviews, Jira-gestütztes Change Management, Unit- und Integrationstests sowie formale Kundenabnahme; Leitprinzipien: Separation of Concerns, Locality of Change und Wiederverwendbarkeit. Fundierter quantitativer Hintergrund (B.Sc. Wirtschaftsmathematik, Schwerpunkt Statistik und Stochastik), der belastbare Evaluierung, Metriken und Modellverständnis untermauert.

TECHNISCHE SKILLS

Data & Lakehouse Engineering Databricks (PySpark · Delta Lake · Unity Catalog · Workflows) · Spark-Internals & Performance-Tuning (Partitionierung · Shuffle · Skew · Join-Strategien) · Medallion-Architektur · dimensionale Modellierung & SCD Type 2 · Data Vault 2.0 · dbt · ETL / ELT · Datenqualität & Reconciliation

GenAI & Agentic AI End-to-End-RAG (Parsing · Chunking · Embeddings · Vector Search · Re-Ranking) · Multi-Agent-Architekturen (Mosaic AI Agent Framework · Genie Spaces · Supervisor Agent) · MCP-Tool-Integration · LLM-Bereitstellung & OpenAI-kompatible Endpoints · Prompt Engineering · LangChain · MLflow

Cloud & Plattform Azure · Azure Data Factory · ADLS Gen2 · Azure AI Foundry · Azure Functions · Azure Key Vault · Microsoft Entra ID / OAuth 2.0 · Azure DevOps · AWS und GCP / BigQuery (produktiv 2020-2022)

Integration & Quellsysteme REST- & SOAP-APIs · konfigurations- und metadatengetriebene Onboarding-Frameworks · strukturierte, semi- und unstrukturierte Quellen (PDF · XML · JSON · CSV · Feeds)

Engineering-Praxis Terraform (IaC) · CI/CD mit Plan-/Apply-Trennung & Approval-Gates · Git & Code-Reviews · Unit- und Integrationstests · Design Patterns (Separation of Concerns · Locality of Change) · SDLC & Change Management (Jira) · Apache Airflow · Docker · FastAPI · Claude Code / GitHub Copilot

Programmiersprachen & Analytics Python · SQL · R · Power BI · Tableau · SAS

Konzeptionell übertragbar Snowflake · Microsoft Fabric · OpenSearch · Exasol / Starburst

BERUFSERFAHRUNG

seit 01/2026

Data & AI Engineer Freelance (mithelfer.ai)

- **Agentic AI & RAG:** Konzeption und Umsetzung von GenAI-Anwendungen für KMU- und Enterprise-Kunden: agentische Decision-Support-Tools und Retrieval-Augmented-Generation-Workflows (RAG), von der Idee bis zum funktionierenden System.
- **Agenten- & Integrationsarbeit:** Umsetzung von AI-Agenten und Automatisierungs-Flows, integriert mit externen Systemen und APIs (FastAPI-Services, containerisiertes Deployment mit Docker), u. a. auf der OpenClaw-Plattform für Automatisierung und AI-Agenten.

Tech Python · RAG · LangChain · FastAPI · Docker · SQL · OpenClaw

02/2026 - 07/2026

Data & AI Engineer Ultra Tendency (Beratung)

- **Lakehouse Engineering:** Konzeption und Aufbau von Ingestion- und Verarbeitungspipelines

in PySpark, Delta Lake und Unity Catalog; Ursachenanalyse und Behebung von Fehlern in produktiven Pipeline-Läufen anhand von Spark-Query-Plan und Job-Metriken; Optimierung des Pipeline-Durchsatzes über gezielte Partitionierung, Reduktion von Shuffle-Aufwand, Behandlung von Datenschiefe und bewusste Wahl der Join-Strategie.

- **Quellintegration & Orchestrierung:** Integration heterogener externer Quellen (REST- und SOAP-Dienste, u. a. Bonitätsdaten) in das zentral verwaltete Lakehouse; Orchestrierung und Scheduling über Databricks Workflows.
- **Datenbasis für GenAI:** Aufbau der Ingestion- und Aufbereitungsschicht (Beschaffung, Parsing und Strukturierung in Delta- und Unity-Catalog-Tabellen), die agentische Assistenten (Genie Spaces) speist und über dasselbe Delta-Muster RAG-Vektorindizes via Databricks Vector Search befüllt; einschließlich Chunking-Strategie, Embedding-Auswahl und Index-Aufbau.
- **Agentic AI auf Databricks:** Aufbau von Multi-Agent-Architekturen (Mosaic AI Agent Framework, Genie Spaces) hinter einem zentral verwalteten Endpoint, bereitgestellt als OpenAI-kompatible REST-APIs (FastAPI), erweitert um MCP-Tools für den Zugriff auf operative Systeme und um einen RAG-Pfad über Databricks Vector Search.
- **Engineering & Change Management:** Git-basierte Entwicklung mit Code-Reviews, Infrastructure as Code (Terraform auf Azure), CI/CD und Jira-gestützten Ticket-Workflows in einer regulierten Unternehmensumgebung.
- **Test & Abnahme:** Unit- und Integrationstests, systematische Ursachenanalyse und formale Kundenabnahme.

Tech Databricks · PySpark · Delta Lake · Unity Catalog · Mosaic AI · Genie Spaces · MCP · Azure AI Foundry · MLflow · Vector Search · Python · SQL · Azure · Terraform · FastAPI · Git · Jira

01/2024 - 07/2025

Data Engineer / Data Analyst CBTW (Beratung)

- **DWH-zu-Lakehouse-Migration:** Vorantreiben der Migration bestehender Data-Warehouse-Workloads nach Databricks; Überführung etablierter ETL-Logik in dbt-Modelle, Notebooks und Jobs (PySpark, SQL).
- **Pipeline Engineering:** Entwicklung und Optimierung von Ingestion- und Transformationspipelines für Multi-Source-Datensätze; Steigerung von Durchsatz und Datenqualität gemeinsam mit den Data-Engineering-Teams.
- **Dashboards & Analytics:** Aufbau und Tuning von Prognosemodellen sowie Entwicklung interaktiver Power-BI- und Tableau-Dashboards für Kunden aus Banking und Versicherung.
- **Delivery-Prozess:** Arbeit in Git-/Jira-gestützter Delivery mit Code-Reviews, Unit-Tests und ticketbasiertem Change Management.

Tech Databricks · PySpark · dbt · Python · SQL · Azure · Azure Data Factory · ADLS Gen2 · Apache Airflow · Power BI · Tableau · Git · Jira · SAS

11/2022 - 01/2024

Data Engineer / Data Analyst Freelance (statistikberatung.eu)

- **Statistikberatung:** Fortgeschrittene quantitative Methoden (Regression, Clustering, Hypothesentests) und statistische Modellierung für Kunden aus Gesundheitswesen, Finanzwesen und Wissenschaft.
- **Data Engineering & Analytics:** Datenaufbereitung, Modellierung und Reporting über heterogene Datensätze hinweg; reproduzierbare Analyse-Workflows in Python und R.

Tech Python · R · SQL · Azure · Power BI

05/2020 - 10/2022

Data Engineer / Data Analyst FocusEconomics

- **Datenpipelines auf BigQuery:** Aufbau und Optimierung von Ingestion- und Transformationsstrecken auf Google BigQuery; Integration großer Mengen an Wirtschafts- und Marktdaten aus heterogenen externen Quellen in eine konsistente, abfragebereite Plattform, mit messbaren Verbesserungen bei Effizienz und Aktualität der Datenbereitstellung.
- **Einführung von Databricks:** Etablierung von Databricks in der bestehenden Datenverarbeitungslandschaft; Aufbau von Pipelines auf AWS und Databricks zur Datenaufnahme und -verarbeitung sowie für Reporting und Analytics, als Grundlage für eine besser skalierbare Plattform.
- **Datenmodellierung & Reporting:** Entwicklung skalierbarer Datenmodelle und Integrationslogik, die die BigQuery-Umgebung an das nachgelagerte Power-BI-Reporting anbinden; Analyse zentraler Wirtschaftstrends und Aufbereitung komplexer Datensätze in klare visuelle Darstellungen für Kunden und interne Teams.
- **Daten- & Code-Qualität:** Verbindliche Standards bei Daten und Code sowie in der Fehleranalyse; zuverlässiger Datenbetrieb trotz Umfang und Häufigkeit der Aktualisierungen.

Tech Google BigQuery · Databricks · dbt · AWS · Python · SQL · Power BI · Tableau

08/2017 - 11/2017

Werkstudent Data Analyst, Controlling tecis Finanzdienstleistungen AG

- **Financial Modeling & Controlling:** Integration von Finanzmodellen und Performance-Kennzahlen in einer regulierten Finanzdienstleistungsumgebung.

AUSGEWÄHLTE PROJEKTE

03/2026 - 07/2026

Handel: Databricks-Lakehouse- und Agentic-AI-Plattform (AI-Center)

Data & AI Engineer

Agentic AI auf Databricks: Konzeption und Deployment eines Databricks Supervisor Agent (Mosaic AI), der mehrere Genie Spaces (Verkaufszahlen nach Region, Filiale und Produktkategorie; Lagerbestände; Lieferanten- und Bestelldaten) hinter einem zentral verwalteten Endpoint orchestriert, bereitgestellt als OpenAI-kompatible REST-API und abgesichert über OAuth 2.0 (Microsoft Entra ID). Erweiterung um MCP-Tools für Zugriff auf operative Systeme, Auslösen von Reports und Abfrage externer APIs, ergänzt um einen RAG-Pfad über Databricks Vector Search (Parsing, Chunking, Embeddings, Re-Ranking) für unstrukturierte Dokumentbestände. LLM-Bereitstellung, Prompt-Flow-Orchestrierung und Model-Monitoring über Azure AI Foundry; Anbindung eines selbst gehosteten Chat-Frontends (LibreChat) über denselben Endpoint, inklusive durchgängiger Identitätsweitergabe: eine eigens gebaute FastAPI-Zwischenschicht reicht das Entra-ID-Token des angemeldeten Mitarbeiters bis nach Databricks durch, sodass Berechtigungen im Backend greifen. Ermöglichte Self-Service-Zugriff auf Unternehmensdaten in natürlicher Sprache über Fachbereiche hinweg und reduzierte den Bedarf an manuellem Ad-hoc-Reporting deutlich.

Lakehouse Infrastructure as Code & Secret Management: Stabilisierung und Weiterentwicklung einer Terraform-verwalteten Lakehouse-Infrastruktur über mehrere Umgebungen in einem regulierten Umfeld. Diagnose und Behebung eines systemischen State-vs.-Runtime-Drifts im Credential- und Secret-Management über Terraform-State, Key Vault und Databricks-Identitäten; Wiederherstellung deterministischer und idempotenter Deployments. Governance gestufter Azure-DevOps-Pipelines mit kontrollierter Plan-/Apply-Trennung und Approval-Gates, um Risiken produktionskritischer Änderungen vor dem Rollout sichtbar zu machen und zu entschärfen.

Externe API-Integration für Bonitätsdaten: Konzeption und Umsetzung einer End-to-End-Ingestion-Pipeline, die einen SOAP-basierten externen Anbieter in ein Cloud-Lakehouse unter Unity-Catalog-Governance integriert (wiederverwendbarer API-Client, serverlose Verarbeitung, orchestriertes Scheduling und strukturierte Aufbereitung in Delta-Tabellen), verifiziert durch einen formalen Abnahmetest mit 17 Testfällen, die der Client beherrschen musste - einschließlich zustandsverändernder Operationen wie Passwortwechsel; der Agent konnte damit nicht nur lesen, sondern auch handeln. Idempotenz des Ladelaufs über einen zusammengesetzten Business-Key im Delta-Merge.

Tech Databricks · Mosaic AI · Genie Spaces · MCP · Vector Search · Delta Lake · Unity Catalog · Azure AI Foundry · Terraform · Azure DevOps Pipelines · Azure Data Factory · Azure Functions · Azure Key Vault · Entra ID / OAuth 2.0 · LibreChat · Python · FastAPI · SQL · Git

03/2025 - 07/2025

Banking: Konfigurationsgetriebenes Onboarding-Framework & automatisierte Reconciliation

Data Engineer

Konzeption und Aufbau einer zentralen Plattform, die heterogene interne und externe Quellsysteme in ein Lakehouse unter Unity-Catalog-Governance auf Azure und Databricks konsolidiert. Entwicklung eines konfigurations- und metadatengetriebenen Onboarding-Frameworks: Neue Quellsysteme werden allein über Konfiguration angebunden, ohne Eingriff in den Pipeline-Code (Separation of Concerns, Locality of Change). Eine zentrale Steuerungsschicht orchestriert die End-to-End-Ingestion über Azure Data Factory und speist eine einheitliche Databricks-Modellierungsschicht. Dort werden Quelldaten entlang der Medallion-Architektur in Delta-Tabellen aufbereitet: Modellierung des integrierten Kernbestands nach Data Vault 2.0 (Hub, Link, Satellite) mit ladehistorisierten Satelliten, darauf aufbauend dimensionale Auswertungsschichten mit SCD-Type-2-Historisierung - Grundlage für das gesamte nachgelagerte Portfolio. Auf dieser Plattform Umsetzung einer vollautomatisierten Abstimmung (Reconciliation) auf Transaktionsebene, die interne Ledger-Buchungen gegen externe Nostro-Kontoauszüge abgleicht: Valuta- und Soll-/Haben-Logik, konfigurierbare Toleranzen, Kategorisierung von Abweichungen sowie Match-/Mismatch-Reporting. Enge Zusammenarbeit mit Finance, Operations und Plattform-Teams, mit Wiederverwendbarkeit, Skalierbarkeit und Datenqualitätskontrollen als Leitprinzipien.

Tech Azure Data Factory · Databricks · PySpark · Delta Lake · Medallion-Architektur · Data Vault 2.0 · SCD Type 2 · Reconciliation · ADLS Gen2 · Python · SQL

01/2025 - 06/2025

Consulting / IT: Databricks Workshop-Entwicklung & Training

Data Engineer / Trainer

Konzeption, Erstellung und Durchführung praxisorientierter Databricks-Workshops und Trainings für Entwicklerteams und Fachanwender. Vermittlung von Best Practices in PySpark, Delta Lake, Unity Catalog und Medallion-Architektur; Erstellung von Workshop-Materialien, Hands-on-Übungen und Referenzimplementierungen. Befähigung der Teilnehmer zur eigenständigen Entwicklung von Datenpipelines und Analytics-Lösungen sowie begleitendes Coaching während

der Umsetzungsphase.

Tech Databricks · PySpark · Delta Lake · Unity Catalog · Medallion-Architektur · Python · SQL · Power BI

12/2024 - 02/2025

Consulting / IT: Skalierbare Power-BI-Architektur mit Databricks-Integration

Data Engineer / Data Analyst

Konzeption einer sicheren, skalierbaren Power-BI-Reporting-Architektur auf Azure mit Databricks-Integration: Datenmodelle und Governance-Konzepte, Dev-/Test-/Prod-Deployment-Strategie, RBAC und Row-Level Security.

Tech Azure · Databricks · Power BI · SQL

05/2024 - 11/2024

Versicherung: Migration der SAS Marketing Automation auf SAS CI 360

Data Engineer / Data Analyst

Migration und Modernisierung von Marketing-Automation-Prozessen auf die cloudbasierte Plattform SAS CI 360: Redesign der Datenmodelle für Cloud-Kompatibilität und Koordination zwischen Marketing und IT.

Tech SAS CI 360 · SAS Base · SQL

02/2024 - 06/2024

Consulting / IT: ESG-Strategie und Dashboard

Data Engineer / Data Analyst

Mitgestaltung einer unternehmensweiten ESG-Strategie und eines zentralen ESG-Dashboards für das Management-Reporting: KPI-Definition, Reporting-Automatisierung und Datenaufbereitung auf Databricks, ausgerichtet auf die CSRD-Anforderungen.

Tech Databricks · Power BI · SQL

01/2024 - 05/2024

Energie: SAS-basierte ETL- und Data-Warehouse-Entwicklung

Data Engineer / Data Analyst

Konzeption und Implementierung von ETL-Prozessen und Data-Warehouse-Lösungen für große operative Datenmengen, orchestriert mit Apache Airflow, inklusive Optimierung der DWH-Strukturen und Sicherung der Datenqualität.

Tech SAS DI · SAS Base · SQL · Apache Airflow · HTML · JavaScript

AUSBILDUNG

10/2015 - 10/2019

B.Sc. Wirtschaftsmathematik Technische Universität Berlin

Schwerpunkt: Statistik & Stochastik.

ZERTIFIZIERUNGEN & WEITERBILDUNG

- Databricks Certified Associate Developer for Apache Spark, in Vorbereitung, Prüfung geplant Q3/2026
- Databricks Certified Data Engineer Professional, in Vorbereitung, Prüfung geplant Q3/2026
- Databricks Certified Data Analyst Associate
- Databricks Fundamentals Accreditation
- SAS Certified Specialist: Base Programming

SPRACHEN

Deutsch C2, Muttersprache · Englisch C1, verhandlungssicher

REFERENZEN

Schriftliche Referenzen auf Anfrage: Ultra Tendency (Arbeitszeugnis) · CBTW (Arbeitszeugnis) · FocusEconomics (Empfehlungsschreiben). Telefonische Referenzen aus früheren Projekten nach Absprache möglich.

Berlin, August 2026 · Elyas El Bastawissy