

# HASSAN RAUF BUTT

+49 176 21444539 | Berlin, Germany

hassanraufbutt21@gmail.com | linkedin.com/in/hassan-rauf-butt | github.com/HassanButt2019

## OBJECTIVE

Data Engineer specializing in scalable data platforms, ETL pipelines, and real-time data processing. Skilled in Python, FastAPI, PostgreSQL, and Docker, with hands-on experience in CI/CD, data modeling, and deploying analytics-ready infrastructure for cross-functional teams.

## EDUCATION

**Master of Artificial Intelligence**, Brandenburg University of Technology  
**Bachelor of Science in Computer Science**, FAST National University

Oct 2023 – Present  
2018 – 2022

## SKILLS

<b>Programming &amp; Frameworks</b>	Python, SQL, Bash, PySpark, FastAPI
<b>Data Engineering</b>	ETL/ELT, Data Modeling (Star/Snowflake, SCD), Batch & Streaming; API Integration
<b>Orchestration/Transform/Processing</b>	Airflow, Dagster, dbt, Apache Spark, DuckDB, Kafka
<b>Databases, Warehousing &amp; Formats</b>	PostgreSQL, MongoDB, Redis, Parquet, Avro
<b>DevOps &amp; CI/CD</b>	Docker, GitHub Actions
<b>Data Quality &amp; Governance</b>	Great Expectations
<b>AI/ML for Data Engineering</b>	Feature Engineering; Time Series Forecasting (XGBoost, LightGBM, Prophet, SARIMA), Data Prep for ML Pipelines
<b>Cloud</b>	AWS (S3, EC2, RDS, SNS, SQS, CodeCommit, Lambda, Cognito), Firebase, Azure, GCP

## EXPERIENCE

<b>Data Engineer</b> MylanTech GmbH	Feb 2025 – Present <i>Munich, Germany</i>
<ul style="list-style-type: none"><li>Built a scalable data platform to ingest, transform, and store 500+ economic indicators from APIs and structured files (Trade Economics, OECD, BLS, IMF), powering analytics and ML pipelines across the product.</li><li>Engineered asynchronous ETL pipelines in FastAPI and Python, achieving 40% faster ingestion via parallel data fetching and background task orchestration.</li><li>Modeled and maintained relational (PostgreSQL) and NoSQL (MongoDB) databases, optimizing schema design and indexing for efficient querying and long-term scalability.</li><li>Standardized heterogeneous data sources, resolving schema mismatches and time alignment issues across 10+ pipelines, ensuring analytics-ready datasets.</li><li>Automated CI/CD workflows using Docker, and GitHub Actions for data pipelines and services, enabling reliable, version-controlled production deployments.</li><li>Prepared feature-rich time series datasets by engineering lag features, applying correlation-based feature selection, and exporting in multiple formats (CSV, JSON, Parquet) for analytics and ML pipelines.</li></ul>	
<b>Student Assistant - Backend Engineer</b> German Research Center for Artificial Intelligence (DFKI)	Apr 2024 – Sep 2024 <i>Bremen, Germany</i>

<ul style="list-style-type: none"><li>Developed scalable backend systems with FastAPI and MongoDB, leveraging asynchronous programming and GridFS to optimize image processing, concurrent workflows, and real-time responsiveness.</li><li>Enhanced AI-driven task scheduling with Google OR-Tools and contributed to open-source tools supporting Reinforcement Learning in robotics.</li></ul>	
---	--