

Rupert Lorenz

Senior AI Data Engineer

✉ blackaigentle120@gmail.com ☎ +49 15510 235616 📍 Frankfurt, Germany 📞 +49 15510 235616

Profile

Senior AI Data Engineer with 9 years of professional expertise in Data Engineering, Machine Learning, Cloud Engineering, and DevOps, distinguished by a strong capacity for rapid learning, innovative thinking, and delivering superior-quality work in environments using Scala, Python, and AWS.

- Demonstrates proven expertise in discerning complex business needs within large-scale data environments and crafting sophisticated models and algorithms tailored to client specifications, using Scala and Spark to handle large volumes of data with high efficiency.
- Highly skilled in designing and implementing bespoke tools and frameworks that enhance automation and scalability, as well as deploying and managing robust data pipelines using AWS services like Glue, Lambda, and Step Functions.
- Expert in fostering and maintaining strong relationships with key stakeholders, consistently achieving successful outcomes through effective communication, strategic collaboration, and deep technical knowledge in data processing with Scala and Python.
- Accomplished in integrating machine learning and data processing workflows into scalable cloud solutions, ensuring optimal performance and reliability across various AWS platforms.
- Adept at using Scala and Python to develop and optimize data transformation processes, enabling faster decision-making and improved operational efficiency in dynamic and demanding data engineering contexts.

Professional Experience

Senior AI Data Engineer, CQUELLE GmbH

08/2020 – 03/2024

Frankfurt, Germany

- As a Senior AI Data Engineer on the GPTLawyer.pro project, I utilized Scala and Python to design, develop, and launch the AI-driven legal advice platform, contributing significantly to a 30% increase in user engagement within the first three months post-launch through effective data processing and analytics.
- Developed real-time streaming pipelines using Apache Kafka and Scala, handling millions of messages from various data sources, enabling real-time monitoring and analytics, and streaming processed data into PostgreSQL.
- Collaborated with stakeholders and led Scrum Teams using Scala and Python to ensure that products, data management, data governance, and data quality met customer satisfaction and aligned with corporate commitments.
- Orchestrated complex data workflows with Apache Airflow and Scala, ensuring smooth execution and efficient processing of large datasets, enhancing task dependencies management, and optimizing data analysis workflows.
- Led the design, testing, and deployment of AI functionalities in 13 projects worth \$4.7M, focusing on Scala-based LLM innovations and LangChain, significantly boosting project value and client satisfaction through advanced data engineering techniques.
- Developed scalable data architectures and maintained high-performance ELT/ETL pipelines in cloud environments using Scala, RedShift, AWS Glue, EMR, and ECR to manage data from various sources and automated pipeline deployment using CI/CD tools like GitHub/GitLab.

Senior Data Engineer, Remoteplatz

09/2017 – 06/2020

Berlin, Germany

- Specialized in advanced ETL processes and data warehousing using Scala and Spark, enhancing system efficiency by 30% and improving data accuracy by 40% across company projects.
- Led a healthcare analytics firm's data system overhaul using Scala and Python, boosting analysis efficiency by 25% through enhanced data quality and optimized processing, demonstrating deep data engineering expertise.
- Built efficient Scala-based data pipelines in healthcare data management, leading to a 30% faster data analysis, crucial for enhancing patient care and treatment outcomes.
- Designed a cutting-edge data architecture for a large-scale NLP project using Scala and Spark to process 5 TB of text data daily, boosting system reliability and scalability.
- Implemented a Biometrics Project using Scala, optical sensors, and pattern recognition algorithms, increasing the company's monthly income by 26% through advanced data processing techniques.
- Implemented data processing and transformation tasks using Apache Spark and Scala, optimizing the performance of distributed computing frameworks in large volume data transformations within AWS.

Machine Learning Engineer | Data Analyst, mindworks GmbH

02/2014 – 08/2017

Hamburg, Germany

- Trained and fine-tuned machine learning models using Scala and Python to detect and classify anomalies in the manufacturing process of 34+ industrial robots, reducing their monthly downtime by 83% and significantly enhancing operational efficiency.
- Contributed to the development of machine learning models using Scala and Python, aiding in a 10% improvement in accuracy, under the mentorship of senior engineers, and leveraging Spark within AWS for scalable data processing.
- Guided the integration of various databases into a single data source using Scala and Python, applying practices and strategies that have been scaled out to other data platforms within AWS environments.
- Partnered directly with the executive team as the first data hire to formulate and report on KPIs across their web properties that received 225M visitors annually, using SQL and Google Sheets, and optimizing data flows with AWS services.
- Dockerized, deployed, and built CI/CD pipelines for several AI products using Git, CircleCI, Trello, and integrated Scala-based Spark jobs to enhance automation and deployment within AWS ecosystems.
- Worked closely with data scientists and analysts to understand their requirements and provided them with clean and reliable data for analysis and modeling using Scala and Python, ensuring optimal use of AWS services like Glue and Lambda for data transformations.

Education

Bachelor of Computer Science, University of Munich

04/2010 – 08/2013

München, Germany

- Ranked in the top 5% of graduating class.

Skills

Programming Languages — Expert

Python, Scala, C++, JavaScript, Go, R, Java, Scala, C#, MATLAB, SQL

Data Science & Machine Learning — Expert

Tensorflow, PyTorch, Scikit-learn, Numpy, Pandas, MLFlow, KubeFlow

Data Visualization — Expert

Matplotlib, Seaborn, Plotly, Dash, Bokeh, D3, ggplot, Tableau, PowerBI

AWS Services — Expert

AWS Glue, AWS Lambda, AWS Step Functions, Amazon S3, Amazon Redshift, AWS Fargate/ECS

Data Engineering — Expert

Apache Airflow, Apache Hadoop, Apache Spark, Apache Beam, Apache Kafka, BigQuery, Snowflake, Redshift

Database Management — Expert

MySQL, PostgreSQL, MongoDB, DynamoDB, Neo4j

DevOps — Expert

Git, DVC, Jenkins, CircleCI, CML, Terraform, Docker, Kubernetes