

A highly skilled Big Data Consultant with over four years of experience in working with big data technologies, primarily Spark and AWS Cloud. Certified as Databricks Certified Associate Developer for Apache Spark 3.0, AWS Solutions Architect – Associate, and AWS Machine Learning – Specialty. - Developed a big data processing engine for large-scale RDF (Resource Description Framework) data during their M.Sc. - Experience in migrating data preparation logic from Alteryx workflows to Spark/Scala scripts, cleaning/processing sensor data by writing PySpark jobs, and developing ETL jobs for online shopping recommendation use cases. - Currently working with a clothing company to create and maintain their data pipelines in Dataiku, optimizing Dataiku workflows and reducing execution time by up to 70%. Key Skills: DBMS (MySQL, MS SQL Server, SPARQL, HBase, Redshift), Development Tools (IntelliJ IDEA, GitHub, Jira, Confluence, Bitbucket, Maven, Eclipse), Operating Systems (Windows, MacOS, Linux), Platforms (AWS, Apache Spark, Dataiku, Databricks, Azure, Snowflake), Programming Languages (Python, Scala, Java, SQL). Fluent in English and have intermediate proficiency in German. Previous Experience: Worked as a Data Engineer at a Clothing Company in Munich, reducing ETL pipeline execution time by up to 70% and developing data pipelines for inventory distribution across different location clusters. Previously worked as a Data Engineer at a Supermarket Chain in Munich, developing an online shopping recommendation system using Databricks, Azure blob storage, and Snowflake. Also worked at a Vehicle Manufacturer in Munich, with unspecified projects. Has experience with Dataiku, PySpark, Python, SQL, Power Query, Databricks, Azure blob storage, Snowflake, and various data engineering tasks such as ETL pipeline development, optimization, and data quality checks. Previously worked on patent landscape analysis, patent monetization, and infringement investigations using MapReduce on the Hadoop framework.