Dmitry Bugakov

Senior Data Engineer |Technical lead Berlin, Germany.

Email: dmitrybugakovdev@gmail.com in LinkedIn G GitHub

ABOUT ME

Data Engineer | Architect with a Computer Science degree and over 10 years of experience in designing data-focused solutions. This includes creating data pipelines, analytical tools, data lakes, and warehouses. I have a keen interest in OpenSource projects and am driven by a passion for leveraging technology to better our world.

- Strong understanding and hands-on experience with Big Data tools such as Hadoop, Spark, and Airflow, along with high-capacity and resilient systems.
- Well-versed in cloud platforms like AWS and GCP.
- Have built streaming solutions using tools such as Kinesis, Kafka, and Pub/Sub combined with Spark streaming.
- Proficient in developing data-centric architectures, ensuring data governance, and maintaining data quality.
- Skilled in navigating SDLC processes, upholding code quality standards, and implementing CI/CD processes with monitoring pipelines.
- Familiar with DataOps and MLOps methods.
- Good at designing systems, starting new projects, and building prototypes.
- Conducted over 100 technical interviews as part of hiring processes.

EXPERIENCE

Deutsche Bank %

Berlin, Germany

Senior Data Engineer |Technical lead

August 2021 - now

Project: Crafted a Data Platform and Data Lake tailored for BI, ensuring swift access to marts with minimal latency.

- Led the revamp of the data platform, minimizing legacy components and transitioning over 100 pipelines to contemporary methodologies.
- Enhanced development processes by integrating Agile methodologies and fortifying CI/CD pipelines, resulting in heightened productivity and expedited deployment.
- o Oversaw SRE and SL3 practices to ensure reliability and optimal performance.

<u>Technologies:</u> Java, Scala, Python, Maven, Hadoop, Clickhouse, Akka, Apache Spark, Pandas, Jupyter, Tableau, Bitbucket, TeamCity etc.

• Grid Dynamics %

Staff Data Engineer |Technical lead

Saint Petersburg, Russian Federation November 2020 - June 2021

Project: Vontier - Developed a combined Data and Data Science platform for enhanced data analysis and model management.

- o Architected data platform solutions, ensuring optimal performance and scalability.
- o Directed data engineering initiatives, leading innovations and ensuring best practices.
- Implemented MLOps and streamlined CI/CD processes, along with A/B test monitoring.

<u>Technologies:</u> Python, PySpark, Pandas, Airflow, Scikit learn, Statsmodels, MLflow, AWS(S3, EMR, Redshift, Kinesis, SageMaker, IAM, KMS, Cloudformation), Jenkins, and more.

Senior Data Engineer | Technical lead

May 2020 - November 2020

Project: HelloFresh, Germany - Spearheaded the redesign and migration of Data Platform and Data Lake to accommodate rapid growth during the pandemic.

- o Architected and executed the data platform, creating an efficient data processing framework.
- o Led the data engineering team, driving innovations and best practices.
- Streamlined development with Agile methodologies and enhanced CI/CD pipelines, boosting productivity and deployment speed.

<u>Technologies:</u> Python, Scala, Spark, Deequ, AWS suite (including EMR, Glue, and Redshift), PostgreSQL, Circleci, Docker, Jira, GitHub, and more.

Senior Data Engineer |Technical lead

July 2019 - May 2020

Project: Kohls - Engineered a Data Platform and Data Lake for the prominent American retailer, Kohl's Corp. (Project concluded amidst the pandemic.)

- o Developed an efficient data processing framework while optimizing GCP interactions to reduce costs.
- o Delivered detailed designs, specifications, and comprehensive documentation.
- o Guided and nurtured team members, emphasizing technical proficiency and fostering a culture of innovation.

<u>Technologies:</u> Scala, Python, Cats, Hadoop, Spark, Airflow, Maven, GCP(Dataflow, Dataproc, Pub/Sub, BigQuery), Vault, Protegrity, Pgp, Zeppelin, Jenkins, Docker, Confluence, Jira, Sonar, Gerrit, Git, etc.

PerfectArt % Moscow, Russian Federation
 Senior Data Engineer | Technical lead
 June 2018 - June 2019

Project: Constructed a Data Platform and Data Lake for DB Schenker, a premier logistics company in Germany.

- Actively contributed to the overarching design of the data platform architecture.
- o Engineered and implemented key parts of data ingestion into the platform.
- Streamlined CI/CD processes and optimized git workflows.

<u>Technologies:</u> Scala, Python, Hadoop, Hive LLAP, Spark, Akka, Slick, AWS (EMR, S3), Kafka, Gitlab, Docker, Jira, among others.

• Sher % Moscow, Russian Federation
Senior Data Engineer July 2016 - May 2018

Factory of Data: A central Data Lake designed to serve all data-centric projects within the bank, acting as the primary source of truth.

- o Developed a data processing framework using Apache Spark and Apache Hadoop.
- o Implemented data processing pipelines with the same technologies.
- o Managed debugging processes, refactored code, and conducted code reviews.

Technologies: Scala, Bash, SBT, Spark, Hadoop, Parquet, Hive, Oozie, Confluence, Jira, Git, Jenkins, among others.

• AT Consulting %

Data Analytics Engineer

Moscow, Russian Federation October 2013 - June 2016

Project: Established a Data Lake for a leading telecom operator in Russia.

- o Development of data processing pipelines using distributed computing technologies
- o Performance optimization and stabilization of a system
- o General debugging, bug fixing and refactoring, code reviewing

<u>Technologies:</u> Java, SQL, Maven, Hadoop, Mapreduce, Hive, Pig, Sqoop, Oozie, Avro, Orc, Oracle, Jira, CentOS (bash), SVN.

EDUCATION

• Voronezh State University & Bachelor's Degree in Mathematics and Computer science (Grade 4.4)

Voronezh, Russian Federation 2010 - 2014