Dmitry Bugakov %

Data Engineering Technical lead/ Data Architect Moscow, Russian Federation



ABOUT ME

Data Engineer/ Architect with a degree in Computer Science and 8+ years of progressive experience designing data-driven solutions including building complex data pipelines, creating analytic frameworks, data lake and data warehouses.

- Deep theoretical knowledge and solid experience with Big Data technology stack (Hadoop, Spark, Airflow), high-load, and fault-tolerant systems.
- Extensive knowledge of using cloud solutions(AWS, GCP).
- Experience building Data Governance, Data Quality, Data Lineage, DataOps.
- Experience building SDLC, code quality, CI/CD, and MLOps processes.
- Experience in system design, development of PoC, and projects from scratch.
- Participation in more than 100 technical interviews as an interviewer.

EXPERIENCE

• Grid Dynamics %

Staff Data Engineer/ Data Architect

Saint Petersburg, Russian Federation
November 2020 - June 2021

Vontier: Data Science platform for working with data and models.

- Design data platform solutions and architecture
- Manage data engineering
- MLOps, CI/CD, release and git flow management
- o A/B testing and monitoring

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

Data Engineering Technical lead

May 2020 - November 2020

Green Chef: Data Lake for the HelloFresh department in Germany.

- o Design data platform solutions and architecture
- Manage data engineering team
- o General debugging, bug fixing and refactoring, code reviewing
- o CI/CD, release and git flow management

<u>Technologies:</u> Python, Scala, Spark, boto3, SQL, AWS(EMR, Glue, Deequ, DMS, S3, Redshift, CloudWatch), PyTest, MongoDB, Parquet, PostgreSQL, Circleci, Docker, Confluence, Jira, GitHub

Data Engineering Technical lead

July 2019 - May 2020

Kohls: Data Lake for American department store retail chain, operated by Kohl's Corp.

- o Design and implement data platform
- o Management of the team of developers
- o Development of functional and implementation designs, specifications, accompanying documentation for the code

Technologies: Scala, Python, Cats, Hadoop, Avro, Spark, Airflow, Maven, GCP(Dataflow, Dataproc, Pub/Sub, BigQuery), ScalaTest, PyTest, Vault, Protegrity, Pgp, Zeppelin, Jenkins, Docker, Confluence, Jira, Sonar, Gerrit, Git

PerfectArt %

Moscow, Russian Federation *June 2018 - June 2019*

Data Engineering Technical lead

DB Schenker: Data Lake for the logistic company in Germany.

- o Development of data processing pipelines using distributed computing technologies
- o Design data lake and data on-boarding process
- o General debugging, bug fixing and refactoring, code reviewing
- o CI/CD, release and git flow management
- Preparation of the training to support customers

<u>Technologies:</u> Scala, Java, Python, SQL, Hadoop, Hive, HBase, Orc, Spark, Airflow, Cats, Slick, Akka-Http, Sbt, AWS(EMR, S3), PostgreSQL, Elasticsearch, Kafka, Filebeat, Logstash, ScalaTest, TestContainers, Gitlab, Docker, MysqlWorkbench, Confluence, Jira, Git

• Sber %

Moscow, Russian Federation November 2017 - May 2018

Machine Learning Engineer

Debt Probability Model: the model helps to assess the likelihood of debt for a client.

- o Contributing in all phases of the development life-cycle and release management
- o General debugging, bug fixing and refactoring, code reviewing
- o A/B testing and monitoring

Technologies: Python2.7, PySpark, Pandas, Numpy, SciPy, Xgboost, Hyperopt, Bash, Confluence, Jira, Git

Senior Data Engineer

July 2016 - November 2017

DataFactory: the Data Lake which is going to be used by all data-dependent projects in the bank. Literally single source of truth for the bank.

- o Development of data processing pipelines using distributed computing technologies
- o Actively involved in overall architecture design, continuous integration adjustment and deployment process
- o Creation proof of possible architecture concept and realization of some of my proposal solutions of these issues
- o Identifying bottlenecks, defects and solutions to these issues
- o General debugging, bug fixing and refactoring, code reviewing

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

AT Consulting %

Moscow, Russian Federation October 2013 - June 2016

Data Engineer

PJSC VimpelCom: Data Lake for the 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

Technologies: Java 8, SQL, Maven, Hadoop, Mapreduce, Hive, Pig, Sqoop, Oozie, Avro, Orc, Oracle, Atlassian Confluence, Jira, CentOS (bash), SVN

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

Voronezh, Russian Federation 2010 - 2014