

# Aleksandr Michailov

E: [mixailow.aleks@gmail.com](mailto:mixailow.aleks@gmail.com) P: +370 696 67 067

LinkedIn: [linkedin.com/in/](https://www.linkedin.com/in/)

GitHub: [github.com/michailow](https://github.com/michailow)

## Profile Summary:

---

Data Engineer with data-driven mindset seeking to leverage my experience in building scalable data infrastructure, integrating and making available structured/unstructured data.

## Experience & Major Projects:

---

### Dashmote, Data Engineer

Feb. 2022 – Oct. 2022

**Spark component development and support.** One of key aspects of our data engineering team is to create and maintain stable data pipelines to provide clean and analytics ready data for business insight generation based on customer needs.

- Owning the development of a new Spark component running on airflow. Additional attention was dedicated to component testing and technical documentation.
- Optimizing old pipeline components. Code refactoring, repo restructuring and adding new features, such as external configurations from parameters files and REST API.
- Performed data exploration and analysis tasks with cloud-based notebook and SQL.

**Airflow pipeline support.** Create new operators, new DAGs. Optimize, and maintain existing pipelines.

**Opening Hours Analysis.** Raw data cleaning, preparation and analysis on Databricks. Creating new ETL scripts.

**Back-end development.** Develop and maintain Restful APIs of internal configuration management tool.

**New members on-boarding.** Helped new colleagues on the technical side by installing software and getting acquainted with the processes taking place in the company.

**Main technologies used:** Python, PySpark, Docker, AWS S3, ECS, EMR, RDS, Airflow.

### Girteka Logistics, Data Migration Specialist

Sep. 2020 – Nov. 2021

**Master Data Migration.** Integrate business data from Microsoft Axapta to SAP.

- Stakeholder handling: collect business requirements, design and implement solutions.
- Python and SQL scripts for Master Data Migration between ERP systems.
- Design solution on Microsoft Azure services using DataBricks.

### IBM Lithuania, Big Data Engineer

Aug. 2019 – Jun. 2020

**Data Pipeline (internship).** Learned to work with modern big data technologies used in big companies and reinforcing knowledge with a practical project.

- The objective of the project was to provide an efficient and scalable solution to migrate data from cloud-based Cloudant DB into Hive DB on Hadoop Cluster.
- Developed an ETL script using PySpark and Big Data Technologies (Hadoop, Hive, and YARN).
- Working and managing remote Linux servers and Big Data Clusters on Docker Images.

**Analytics platform.** Data preparation for chat-bots analytics platform.

- Collaborations with the IBM Research team and providing them with clear and organized data from various sources, like MongoDB, Cloudant, MySQL, Web Requests.
- Developing ETLs scripts with Python Data Stack (Pandas and SQL).
- Improved data quality by analyzing and testing current data pipelines.

## Other projects:

---

- Participated in Online Store Catalog Analysis Automation, July 2020. I found this job through networking. Had studied tools and methods for Data Scrapping and got experience in creating data pipelines using a component of the Microsoft SQL Server database software.

## Key skills:

---

- Python programming.
- ETL script writing using Python Data Stack (NumPy, Pandas, Jupyter Notebook, Matplotlib) and PySpark.
- Notebooks: Jupyter Notebook, Zeppelin, Databricks
- SQL and NoSQL databases (MySQL, PostgreSQL, MSSQL, MongoDB, Cloudant).
- Big Data technologies: Hadoop, Hive, Spark, YARN.
- Workflow management: Apache Airflow.
- Stream data processing: Apache Kafka.
- Cloud technologies: AWS & Azure stacks.
- DevOps tools: Git, Docker, Linux, Jenkins, Jira.
- Agile mindset. Experience with scrum methodology.
- **Languages:** English, Russian, Lithuanian.

## Education:

---

### Computing Science Bachelor Degree, 2015 – 2019

National Research University of Information Technologies, Mechanics and Optics (ITMO University), Saint Petersburg, Russia

Courses: higher mathematics, and physics. Computer science including Python, C#, Assembler, C, C++, as well as MySQL, MongoDB and Data Analysis. Joined the machine learning club during studies.

University had a wide variety of extra-curricular activities, and I actively participated in them. I put a lot of effort into developing my social skills. In the first year, I worked as a volunteer, and to further my growth, also took part as an organizer of events and a curator of first-year students, helping them adapt to university life.

Thesis: Accelerating Data Analysis on a Graphical Processor.

### Maturity Certificate, 2004 – 2015

Gerosios Vilties school, Visaginas, Lithuania

## About me:

---

I have a great passion for technology. I believe in the power of data. Prefer to work in a data-driven organization with like-minded tech-savvy people. I'm eager to learn and to grow.

Would like to find an environment where I can take ownership and responsibilities with proper support.