GESER DUGAROV, Ph.D. | Software Engineer, Big Data Engineer

email: geserdugarov@gmail.com

profiles: LinkedIn, GitHub

SUMMARY

Software Engineer developing core functionality on a data lakehouse platform to extract value from petabyte-scale data. Hands-on experience with Apache Spark and Apache Flink, with a focus on Hudi-Spark and Hudi-Flink integrations. Open-source contributor to Apache Hudi, focusing on streaming performance and solution usability.
Extensive experience in research and data analysis; PhD. Strong interest in big data and distributed systems. Personal mission: "Living a balanced life. Helping professionals work smarter, not harder, by creating automated systems for their routines."

TECHNICAL SKILLS

Java, Python, Maven, PostgreSQL, Docker, Hadoop Ecosystem

WORK EXPERIENCE

May 23 – Current (2.5+ yrs) **Software Engineer / Big Data Engineer**, **Huawei Cloud**

Development of core functionality for big data processing on enterprise-level scalable clusters.

- Provided a simplified configuration system utilizing commonly used presets to overcome the complexity of managing hundreds of parameters.
- Improved performance of Flink stream writing, decreasing processing time by 2x.
- Implemented partition-level TTL, enabling customers to automate cloud storage cost management with coarse granularity.

Jan 24 – Current (2+ yrs) **Apache Hudi Contributor**, **The Apache Software Foundation**

Apache Hudi is a data lakehouse platform that brings database functionality to data lakes and enables incremental processing for low-latency analytics.

- Optimized serialization and deserialization of records in Flink stream writing to Hudi table, resulting in a 30% increase in processing speed and 2x reduction in memory usage (design doc, main changes). Released in Hudi 1.0.2.
- Implemented 4 local optimizations ([1], [2], [3], [4]) resulting in a 10% increase in processing speed and 30% reduction in garbage collection overhead. Released in Hudi 1.0.1.
- Contributed 40+ merged pull requests.

Feb 22 - May 23 (1+ yr) **Software Engineer / ML Engineer**, **Digital Research** (computer vision startup)

- Designed and implemented an event-based architecture for a system for trucks monitoring. Developed server-side image processing handling ~20,000 images per day. In production, the system reduced fleet idle time by 12%.
- Built a customer-facing web UI featuring reports and data visualizations. Also developed an internal web UI for system monitoring.

EDUCATION

PhD, Geophysics, Trofimuk Institute of Petroleum Geology and Geophysics SB RAS

MSc, Computational and Applied Mathematics, Novosibirsk State University

Recent version of CV