|  |  |
| --- | --- |
| **GESER DUGAROV, Ph.D. | Java Developer, Big Data Engineer** | |
| email:  profiles: | [geserdugarov@gmail.com](mailto:geserdugarov@gmail.com)  [LinkedIn](https://www.linkedin.com/in/geserdugarov/), [GitHub](https://github.com/geserdugarov) |

SUMMARY

Software Engineer developing core functionality in a Data Lakehouse platform to extract value from PB-scale data. Also contributing to open source, [Apache Hudi project](https://github.com/apache/hudi/pulls?q=is%3Apr+author%3Ageserdugarov), with a focus on performance and usability of solution. Wide experience in research and data analysis, PhD.

Passionate about Big Data and Distributed Systems. Personal mission: "Living a balanced life. Helping professionals to work smarter, not harder by creating automatic systems for their routine."

TECHNICAL SKILLS

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

WORK EXPERIENCE

|  |  |
| --- | --- |
| May 23 - Current  (2+ yrs) | **Java Developer / Big Data Engineer**,  [**Huawei Cloud**](https://www.huaweicloud.com/intl/en-us/product/mrs.html) |

• Development of core functionality of a Data Lakehouse platform for Big Data processing on enterprise-level scalable clusters.

• Future Star Award (2024).

|  |  |
| --- | --- |
| Jan 24 - Current  (1+ yr) | **Apache Hudi Contributor**,  **The Apache Software Foundation** |

[Apache Hudi](https://hudi.apache.org/) 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 data stream records in Flink stream writing, resulting in a 30% increase in processing speed and 2x reduction in memory usage ([design doc](https://github.com/apache/hudi/pull/12697), [main changes](https://github.com/apache/hudi/pull/12796), [umbrella ticket](https://issues.apache.org/jira/browse/HUDI-8920)). Released in [Apache Hudi 1.0.2](https://github.com/apache/hudi/releases/tag/release-1.0.2).

• Implemented 4 local optimizations ([[1]](https://github.com/apache/hudi/pull/12054), [[2]](https://github.com/apache/hudi/pull/12104), [[3]](https://github.com/apache/hudi/pull/12113), [[4]](https://github.com/apache/hudi/pull/12120)) in Flink stream writing, resulting in a 10% increase in processing speed and 30% reduction in garbage collection overhead. Released in [Apache Hudi 1.0.1](https://github.com/apache/hudi/releases/tag/release-1.0.1).

• Contributed 40+ [merged pull requests](https://github.com/apache/hudi/pulls?q=is%3Apr+author%3Ageserdugarov+is%3Amerged).

|  |  |
| --- | --- |
| Feb 22 - May 23 (1+ yr) | **Python Developer**,  **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](https://geserdugarov.github.io/file/CV_Geser_Dugarov.pdf)