# Big Data Processing and Analytics Inside DBMS

Mikhail Zymbler(✉)[0000−0001−7491−8656], Sachin Kumar,
Yana Kraeva, Alexander Grents, and Anastasiya Perkova

South Ural State University, Chelyabinsk, Russia
mzym@susu.ru, sachinagnihotri16@gmail.com, kraevaya@susu.ru,
grentsav@susu.ru, perkovaai@susu.ru

In the era of Big Data, there are two basic challenges for humans: how to effectively manipulate and analyze huge amounts of data. Currently, relational DBMSs remain the most popular tool for processing large tables in various data intensive domains, despite the widespread use of numerous NoSQL systems. At the same time, most of modern tools for mining the large data sets are non-DBMS and based on the MapReduce paradigm. If we consider DBMS only as a fast and reliable data repository, we get significant overhead for export large data volumes outside a DBMS, changing data format, and import results of analysis back into DBMS. That is why integration of data mining methods with relational DBMS is a topical issue.

There exist parallel DBMSs that can efficiently process transactions and SQL queries on very large databases. Such DBMSs could be a subject for integration of data mining methods but they are expensive and oriented to custom hardware that is difficult to expand. Open-source DBMSs are now being a reliable alternative to commercial DBMSs but there is a lack of open-source parallel DBMSs since the development of such software is rather expensive and takes a lot of time.

In the talk, we will consider an approach to deal with the problems described above. A parallel DBMS can be developed not from scratch but by small-scale modifications of the original codes of an open-source serial DBMS to encapsulate parallelism. Large- and small-scale data mining problems can be solved inside such a parallel DBMS.

## Acknowledgments

This work was financially supported by the Russian Foundation for Basic Research (grant No. 17-07-00463) and by the Ministry of Science and Higher Education of the Russian Federation (government orders 2.7905.2017/8.9 and 14.578.21.0265).