

Informatika i sistemy upravleniya. – 2019. – No. 1(59). – P. 3-17.

Grigorev Yu.A. (grigorev@bmstu.ru), **Proletarskaya V.A.**

Bauman Moscow State Technical University

MODEL OF QUERIES EXECUTION PROCESSES TO DATA WAREHOUSE ON THE PARALLEL COMPUTING PLATFORM SPARK

The execution analysis of SQL queries Q3, Q17 from the TPC-H test in the Spark environment has been performed. A mathematical model based on the analysis of these processes has been developed in order to estimate the time needed to execute queries to the data warehouse for the method of Bloom Filter Cascade Application (BFCA). Based on the results of the full-scale experiments, the parameters of the developed model have been calibrated and its adequacy has been analyzed.

Keywords: SQL, Apache Spark, Bloom filter, TPC-H, Big Data, modeling.

DOI:10.22250/isu.2019.59.3-17

For citation:

Grigorev Yu.A., Proletarskaya V.A. MODEL OF QUERIES EXECUTION PROCESSES TO DATA WAREHOUSE ON THE PARALLEL COMPUTING PLATFORM SPARK // *Informatika i sistemy upravleniya.* – 2019. – No. 1(59). – P. 3-17.