

**Informatika i sistemy upravleniya. – 2017. – No. 1(51). – P. 3-14.**

**Grigorev Yu.A.** ([grigorev@bmstu.ru](mailto:grigorev@bmstu.ru)), **Ermakov E.Yu., Proletarskaya V.A.**  
Bauman Moscow State Technical University

**ACCESS METHOD TO THE WAREHOUSE USING SPARK TECHNOLOGY WITH CASCADING BLOOM FILTER**

This paper is about a new method for execution of SQL-queries in Apache Spark, the environment for parallel computing. It includes a query view of the original request in the form of several sub-queries, development of connection graphic chart and subqueries conversion graphic chart. It also includes compounds identification where it is necessary to use Bloom filter and Spark language graph representation. Taking query Q3 test TPC-H we conducted full-scale experiments confirming effectiveness of the developed method in comparison with the Hive method.

**Keywords:** SQL query, the Spark platform, Bloom filter, TPC-H test, "snowflake" scheme, "star" scheme, Hive, SQLContext, query execution time, performance.

**DOI: 10.22250/isu.2017.51.3-14**

*For citation:*

**Grigorev Yu.A., Ermakov E.Yu., Proletarskaya V.A.** ACCESS METHOD TO THE WAREHOUSE USING SPARK TECHNOLOGY WITH CASCADING BLOOM FILTER // *Informatika i sistemy upravleniya.* – 2017. – No. 1(51). – P. 3-14.