processed data with a SQL query. Thus, the main scenario involves unifying heterogeneous data by converting them to JSON documents and processing them using **Spark SQL**. Other scenarios are also justified, if we take into account the diversity of the source data. For example, the **Spark SQL** module allows direct query to relational databases without their conversion to the JSON format. On the other hand, you can provide access to JSON documents by collecting them in a file system using other Big Data tools. The first and main feature of JSON data collection based on ontological models terms – the unambiguous interpretation of the content and type of data. In this case users and external programs can freely work with data, because the ontology term, mapped to a key or value in the body of the JSON file, has available and accepted definitions and properties. For example, links to various types of files (graphics, multimedia, exe-files, etc.) can be described adequately and functionally using keys-terms from ontologies describing data formats. The second feature, as it is not strange, is the possibility of including in the exchange of such data sources that do not allow active access or changes due to various reasons. Then uploading the data to an external JSON file solves this problem, providing independent data storage and their full description via ontological models.

The listed technologies, supported by **Apache Spark**, provide unlimited productivity and variety of opportunities to handle complex data, which include data on the properties of substances, including traditional materials and nanostructures.

References

[1] Ataeva, O.M., Erkimbaev, A.O., Zitserman, V.Yu. et al.: Ontological Modeling as a Means of Integration Data on Substances Thermophysical Properties. Proc. of 15th All-Russian Science Conference "Electronic Libraries: Advanced Approaches and Technologies, Electronic Collections" – RCDL-2013, Yaroslavl, Russia, October 14–17, 2013. http://rcdl.ru/doc/2013/paper/s1_3.pdf

[2] Introduction to JSON. http://json.org/json-ru.html

[3] 3Vs (volume, variety and velocity), definition from TechTarget Network. http://whatis.techtarget.com/definition/3Vs

[4] Apache Hadoop. https://hadoop.apache.org/

[5] Apache Spark. http://spark.apache.org/docs/

[6] ChemSpider. www.chemspider.com

[7] Frenkel, M., Chirico, R.D., Diky V. et al.: XML-based IUPAC Standard for Experimental, Predicted, and Critically Evaluated Thermodynamic Property Data Storage and Capture (ThermoML) (IUPAC Recommendations 2006). Pure Appl. Chem., 78 (3), pp. 541-612 (2006)

[8] Hall, S.R, McMahon, B.: The Implementation and Evolution of STAR/CIF Ontologies: Interoperability and Preservation of Structured Data. Data Science J., 15 (3), pp. 1-15 (2016). doi: http://dx.doi.org/10.5334/dsj-2016-003

[9] Kaufman, J.G., Begley, E.F.: MatML. A Data Interchange Markup Language. Advanced Materials & Processes/November, pp. 35-36 (2003)

[10] Michel, K., Meredig, B.: Beyond Bulk Single Crystals: A Data Format for all Materials Structure-property-processing Relationships. MRS Bulletin. 41 (8), pp. 617-623 (2016)

[11] Ontobee: A Linked Data Server Designed for Ontologies. www.ontobee.org