

ABOUT THE CONCEPT OF BUILDING A DISTRIBUTED ECOSYSTEM FOR STORING AND PROCESSING DATA WITHIN THE FRAMEWORK OF AN IT SOLUTIONS DEVELOPMENT PLATFORM

G. Dik^{1,*}, *A. Dik*^{1,**}, *A. Bogdanov*²,
*T. Avtomonova*¹, *V. Yurov*²

¹ Saint Petersburg State University, Saint Petersburg, Russia

² Scientific and Production Association “RusBITech”, Saint Petersburg, Russia

We propose the concept of creating a distributed industry platform (DIP) to improve the quality and accelerate developments in the field of IT technologies. The role of database management systems for building a system of data collection, storage and processing in the construction of DIP, where a large amount of heterogeneous data is collected within a single platform, is considered. By creating a local development environment within the platform and deploying the necessary software, developers have the opportunity to develop in the data ecosystem (EDS). In addition, distributed datasets can be created from EDS, and then data analysis systems integrated into the platform can be used. This is especially important when using artificial intelligence tools, where large amounts of raw and prepared data are required for training and testing neural network-based models.

Предложена концепция создания распределенной отраслевой платформы (РОП) для повышения качества и ускорения разработок в области ИТ-технологий. Рассмотрена роль систем управления базами данных для построения системы сбора, хранения и обработки данных при построении РОП, где в рамках единой платформы собирается большой объем разнородных данных. Создав локальную среду разработки внутри платформы и развернув необходимое программное обеспечение, разработчики имеют возможность вести разработку в экосистеме данных (ЭСД). Кроме того, из ЭСД можно создавать распределенные наборы данных, а затем использовать интегрированные в платформу системы анализа данных. Это особенно важно при использовании инструментов искусственного интеллекта, где для обучения и тестирования моделей на основе нейронных сетей требуются большие объемы необработанных и подготовленных данных.

PACS: 44.25.+f; 44.90.+c

* E-mail: g.dick@systechnologies.ru

** E-mail: a.dik@spbu.ru