СЕКЦІЯ 15. ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА ІНТЕЛЕКТУАЛЬНА ВЛАСНІСТЬ

УДК 004.62

А.С. БАБИЧ, НТУ «ХПИ», Харьков, Украина
Е.П. ЧЕРНЫХ, канд. физ.-мат. наук, доц., НТУ «ХПИ», Харьков, Украина

Исследование подходов для обработки больших данных в коммерческих приложениях

Технологии передачи больших данных не только вносят важный вклад, но и представляют собой уникальные вызовы международному развитию. Применение больших данных предлагает экономически эффективные возможности для совершенствования процессов принятия решений в разнообразных областях, таких как здравоохранение, занятость, экономическая продуктивность, преступность, безопасность, управления ресурсами и т.п.

В качестве примеров источников возникновения больших данных могут выступать непрерывно поступающие данные с измерительных устройств, события от радиочастотных идентификаторов, потоки сообщений из социальных сетей, метеорологические данные, данные дистанционного зондирования Земли, потоки данных о местонахождении абонентов сетей сотовой связи, устройств аудио и видео регистрации. Ожидается, что развитие и начало широкого использования этих источников инициирует проникновение технологий больших данных как в научно-исследовательскую деятельность, так и в коммерческий сектор и сферу государственного управления.

Большие данные требуют исключительных технологий и подходов для эффективной обработки. Эти технологии включают в себя краудсорсинг, синтез данных и интеграции, генетические алгоритмы, машинное обучение, обработка естественного языка, обработка сигналов, моделирование, анализ временных рядов и визуализации. Дополнительные технологии, применяющиеся для больших данных, включают распределённые файловые системы, распределённые базы данных, инфраструктуры на основе облако, средства массово-параллельной обработки неопределённо структурированных данных, прежде всего, решениями категории NoSQL, алгоритмами MapReduce, программные каркасы и библиотеки проекта Hadoop.

Существует большое количество разрабатываемых решений для обработки больших данных в разнообразных областях применения, однако все эти решения имеют свои достоинства и недостатки. Таким образом, подходя к решению проблемы больших данных необходимо определить ключевые качества разрабатываемой системы для получения максимально качественного программного продукта.