

АНАЛИЗ РАЗВИТИЯ БОЛЬШИХ ДАННЫХ

Big Data (большие данные) – методы анализа массивов данных большого объёма и разнородного характера (промышленные, экономические, организационные).

Источники больших данных – показания приборов и датчиков, архивы, базы данных, интернет вещей (IoT) – сеть объектов реального и виртуального мира, подключённых к интернету и способных обмениваться данными. Такие данные отличаются рядом характеристик:

Объем. Огромные «объемы» данных, которые организации получают из бизнес-транзакций, интеллектуальных (IoT) устройств, промышленного оборудования, и т. д., нужно где-то хранить.

Скорость. Подразумевает необходимость высокоскоростной обработки данных в реальном времени.

Разнообразие. Разнородный характер больших данных проявляется в их форматах: неструктурированные текстовые, видео- и аудиофайлы, а также полуструктурированная информация.

Большие данные могут быть обработаны в пакетном, потоковом и интерактивном режиме.

Пакетный режим применяется при наличии данных в облачном хранилище и служит для агрегирования данных и построения аналитического отчета на их основе.

Потоковый режим используется для поддержки и автоматизации принятия решений в режиме реального или приближенного к реальному времени.

Интерактивный режим представляет возможность фильтрации, комбинирования, группировки и проведения арифметических операций над данными в режиме реального времени. Это позволяет находить скрытые закономерности в данных.

Таким образом, обработка больших данных необходима для принятия решений по управлению производством (фирмой) на основе прогнозируемых моделей, которые строит искусственный интеллект.

Недостаток Big Data в основном связан с тем, что при их анализе используются сложные алгоритмы, которые нельзя оценить и проверить – неясно какие данные были учтены, почему и как сделан вывод. Кроме того, сбор, хранение и обработка больших объемов информации разнородного характера создает много возможностей для утечек и нарушения конфиденциальности.