



Доклад

на тему: (Большие данные)



Большие данные — совокупность подходов, инструментов и методов обработки структурированных и неструктурированных данных огромных объёмов и значительного многообразия для получения воспринимаемых человеком результатов, эффективных в условиях непрерывного прироста, распределения по многочисленным узлам вычислительной сети, альтернативных традиционным системам управления базами данных и решениям класса.

- NoSQL

NoSQL в информатике — термин, обозначающий ряд подходов, направленных на реализацию хранилищ баз данных, имеющих существенные отличия от моделей, используемых в традиционных реляционных СУБД с доступом к данным средствами языка SQL.

- MapReduce

MapReduce — модель распределённых вычислений, представленная компанией Google, используемая для параллельных вычислений над очень большими наборами данных в компьютерных кластерах.

- Hadoop

Hadoop — проект фонда Apache Software Foundation, свободно распространяемый набор утилит, библиотек и фреймворк для разработки и выполнения распределённых программ, работающих на кластерах из сотен и тысяч узлов.

Введение термина «большие данные» относят к Клиффорду Линчу, редактору журнала Nature, 3 сентября 2008 года



В 2011 году Gartner (исследовательская и консалтинговая компания, специализирующаяся на рынках информационных технологий.) отмечает большие данные как тренд номер два в информационно-технологической инфраструктуре (после виртуализации).

Gartner®

Существуют разные определения больших данных, но большинство из них базируется на концепции «трех V» больших данных:

- **Объем (Volume)**
- **Разнообразие (Variety)**
- **Скорость (Velocity)**

В большинстве случаев работа с большими данными подразумевает стандартный рабочий процесс: от сбора необработанных данных и до получения пригодной для использования информации.

- Сбор. Сбор необработанных данных
- Хранение. Любая платформа для работы с большими данными должна включать надежный, безопасный и масштабируемый репозиторий для хранения данных как до обработки, так и после таковой.
- Обработка и анализ достигается за счет сортировки, агрегации, объединения или применения специальных расширенных функций и алгоритмов
 - Визуализация и использование. Основная цель работы с большими данными – это получение на их основании ценных аналитических выводов для практического применения.