

Распределенные системы и технологии.

# Приложения в распределенных системах

Дмитрий Юрьевич Чалый  
декан факультета ИВТ,  
зав. кафедрой информационных и сетевых технологий



22 мая 2016 г.

# MapReduce: расчет на кластере

- Hadoop (HDFS, отказоустойчивость, мониторинг) — технологическая платформа;
- Map: функция для преобразования индивидуальных значений с результатом (ключ, значение);
- Reduce: функция для комбинирования значений с одинаковым ключом;
- вы реализуете набор функций и указываете в каком порядке они должны быть выполнены.



- Фреймворк для потоковой обработки данных в реальном времени;
- Stream — последовательность кортежей;
- Spout — источник потока (Stream);
- Bolt — преобразование потока (функция, фильтрация, агрегирование и т.д.);
- топология — сеть, в которой объединяются экземпляры Spout и Bolt;
- все работает на кластере.



- C++ API;
- хранение вершин и ребер на тех узлах, где происходит обработка;
- если необходимо, данные о вершинах и ребрах пересылаются на другие узлы.

