

Архитектура программных  
систем, лекция 3.  
Распределённые системы

Пожидаев М. С.

28 марта 2023 г.

# Распределённые системы

## *Ключевые особенности обобщённого понятия*

Под распределёнными системами понимается класс вычислительных систем, которые используют для хранения и обработки информации динамически изменяемое множество узлов.

### Не одноранговая архитектура!

Распределённые системы имеют узел-контроллер, которого нет в одноранговых системах.

### Не микросервисная архитектура!

Вычислительная модель в распределённых системах обычно однородная и больше ориентирована на однотипные операции сортировки и поиска.

# Составные части

## *Компоненты и типы распределённых вычислений*

Распределённая система состоит из:

- ▶ управляющего узла-контроллера, отказ которого приводит к неработоспособности системы;
- ▶ множества узлов-вычислителей, отдельные отказы которых не приводят к неработоспособности системы.

Типы распределённых вычислений:

- ▶ вычисления с сохранением промежуточных данных на диск;
- ▶ вычисления в памяти (in memory).

# Плюсы

## *Достоинства распределённых систем*

1. Способность обрабатывать огромные массивы слабоструктурированных данных.
2. Низкая чувствительность к отказу отдельного вычислительного узла, кроме контролирующих.
3. Простота решения задач сортировки и поиска информации.
4. Возможность географического распределения данных.

# Минусы

## *Недостатки распределённых систем*

1. Наличие критического узла-контроллера., нарушение работы которого приводит к утрате всей системы.
2. Высокая потребность в вычислительных мощностях из-за многократного дублирования хранимой информации.
3. Ограниченность вычислительных задач, выполняемых с хранимыми данными.

# Apache Hadoop

## Платформа распределённых вычислений

- ▶ *HDFS*: распределённая файловая система для хранения сверхбольших массивов данных. Не поддерживает изменения файлов.
- ▶ *MapReduce*: фреймворк для распределённых вычислений.

# MapReduce

## *Порядок вычислений*

Рассмотрим алгоритм подсчёта всех слов в тексте с использованием MapReduce:

1. Map: для каждого слова в тексте формируется пара ключ-значение, в которой в ключ записывается слово, а в значение единица.
2. Получившиеся пары группируются в блоки с одинаковым ключом.
3. Reduce: каждый блок редуцируется до новой пары, состоящей из слова и количества исходных пар.

# Apache Spark

## *Распределённые резидентные вычисления*

1. Способна обрабатывать слабоструктурированные данные.
2. Способна работать в потоковом режиме.
3. Предлагает отклик заметно лучше, чем в обычном MapReduce.
4. Поддерживает некоторые алгоритмы машинного обучения.

# Спасибо за внимание!

Всё о курсе: <https://marigostra.ru/materials/Architecture.html>

E-mail: [msp@luwrain.org](mailto:msp@luwrain.org)

Канал в Телеграм: @MarigostraRu