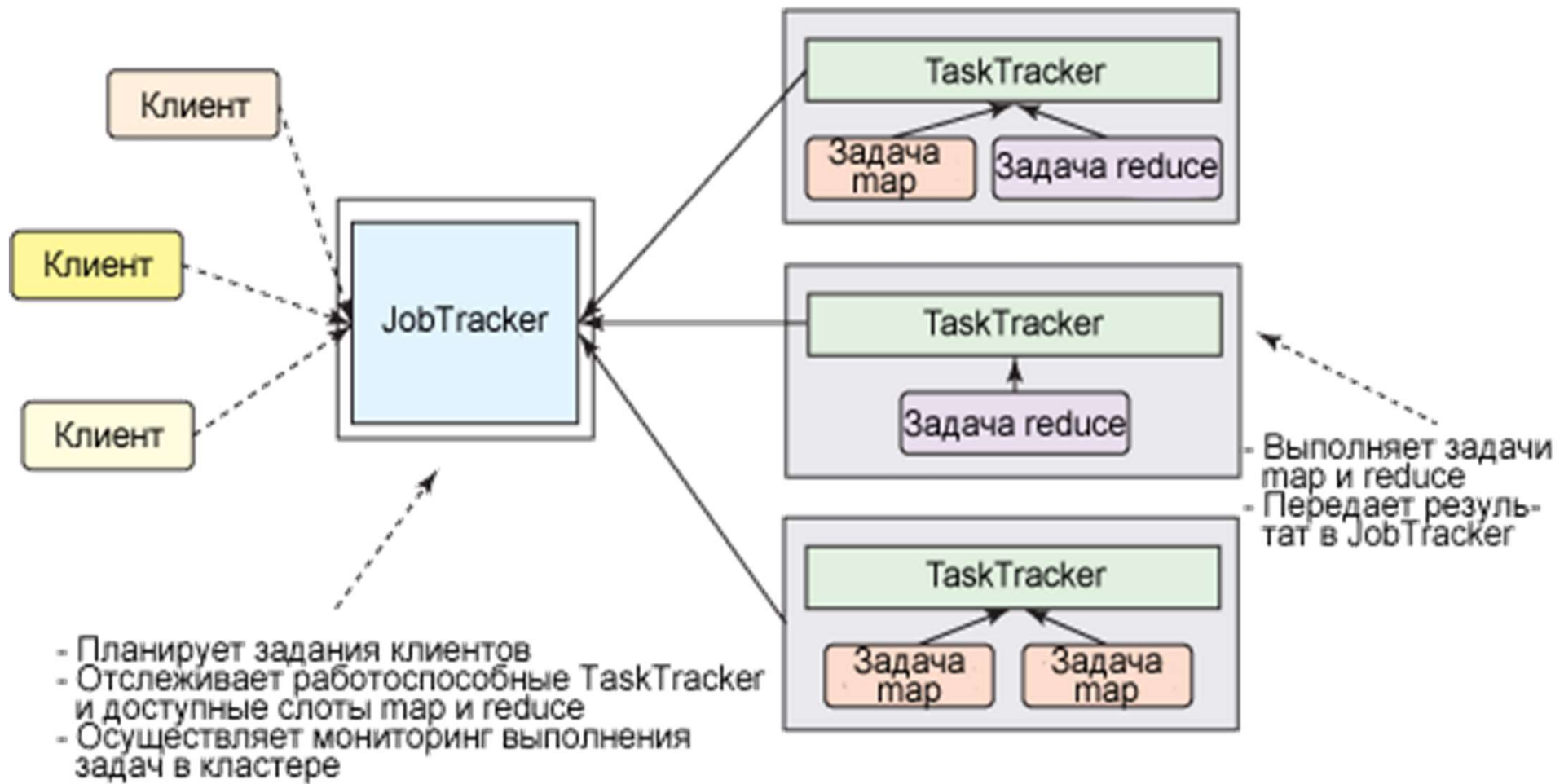


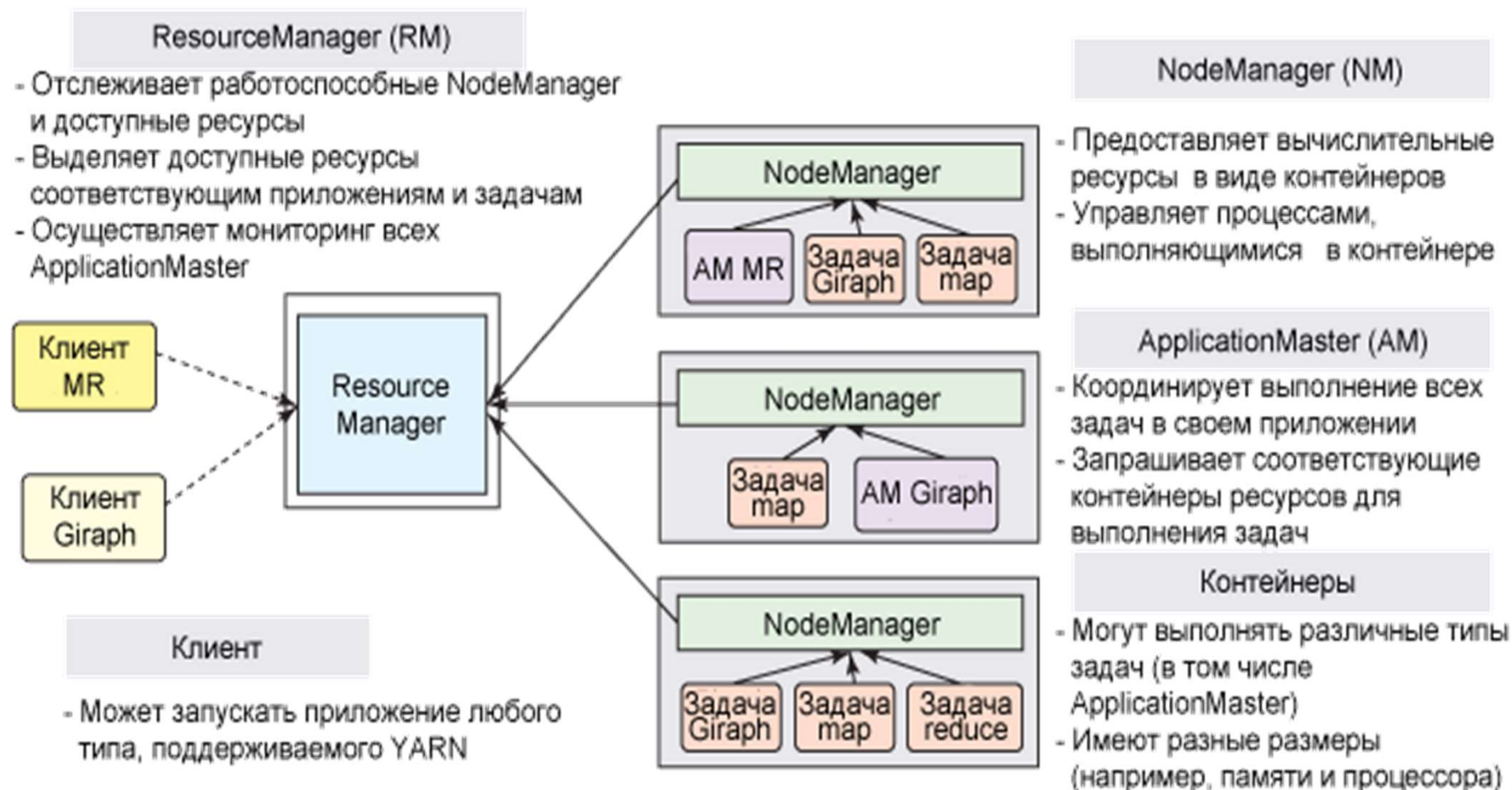
Лекция. Программная платформа Hadoop. Планировщик ресурсов YARN

YARN (*Yet Another Resource Negotiator* — «ещё один ресурсный посредник») — модуль, отвечающий за управление ресурсами кластеров и планирование заданий. Если в предыдущих выпусках эта функция была интегрирована в модуль [MapReduce](#), где была реализована единым компонентом (*JobTracker*), то в YARN функционирует логически самостоятельный [демон](#) — планировщик ресурсов (*ResourceManager*), абстрагирующий все вычислительные ресурсы кластера и управляющий их предоставлением приложениям распределённой обработки. Работать под управлением YARN могут как MapReduce-программы, так и любые другие распределённые приложения, поддерживающие соответствующие программные интерфейсы; YARN обеспечивает возможность параллельного выполнения нескольких различных задач в рамках кластера и их изоляцию. Разработчику распределённого приложения необходимо реализовать специальный класс управления приложением (*ApplicationMaster*), который отвечает за координацию заданий в рамках тех ресурсов, которые предоставит планировщик ресурсов; планировщик ресурсов же отвечает за создание экземпляров класса управления приложением и взаимодействия с ним через соответствующий сетевой протокол.

Классическая версия Apache Hadoop (MRv1)



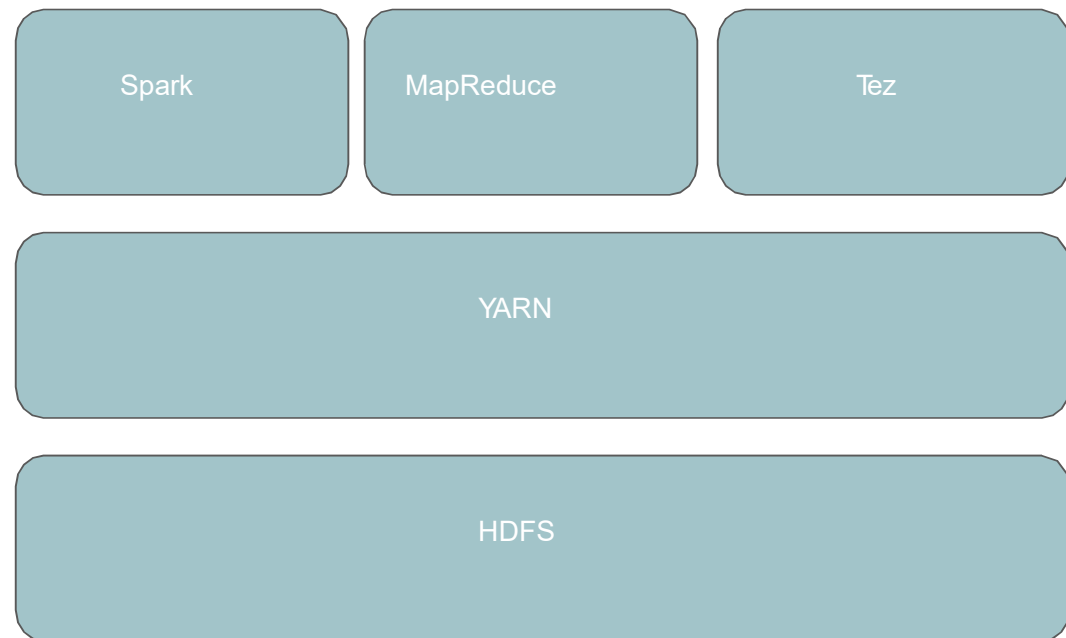
Архитектура YARN



Управление ресурсами: YARN

YARN (“Yet Another Resource Manager”) - утилита, которая:

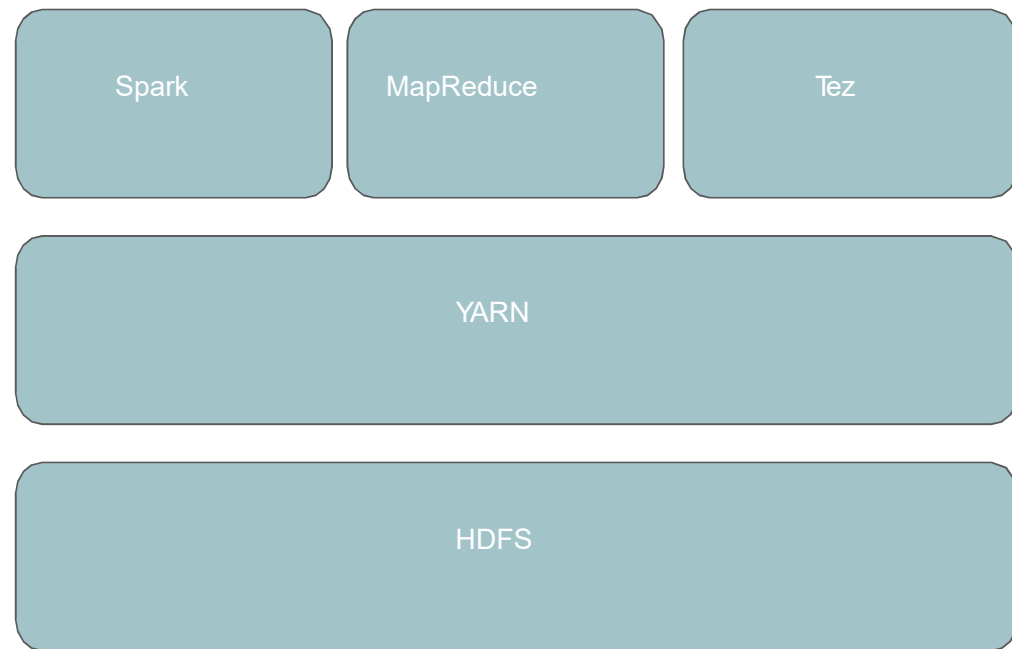
1. Распределяет вычислительные ресурсы кластера (память и CPU) приложениям обработки данных



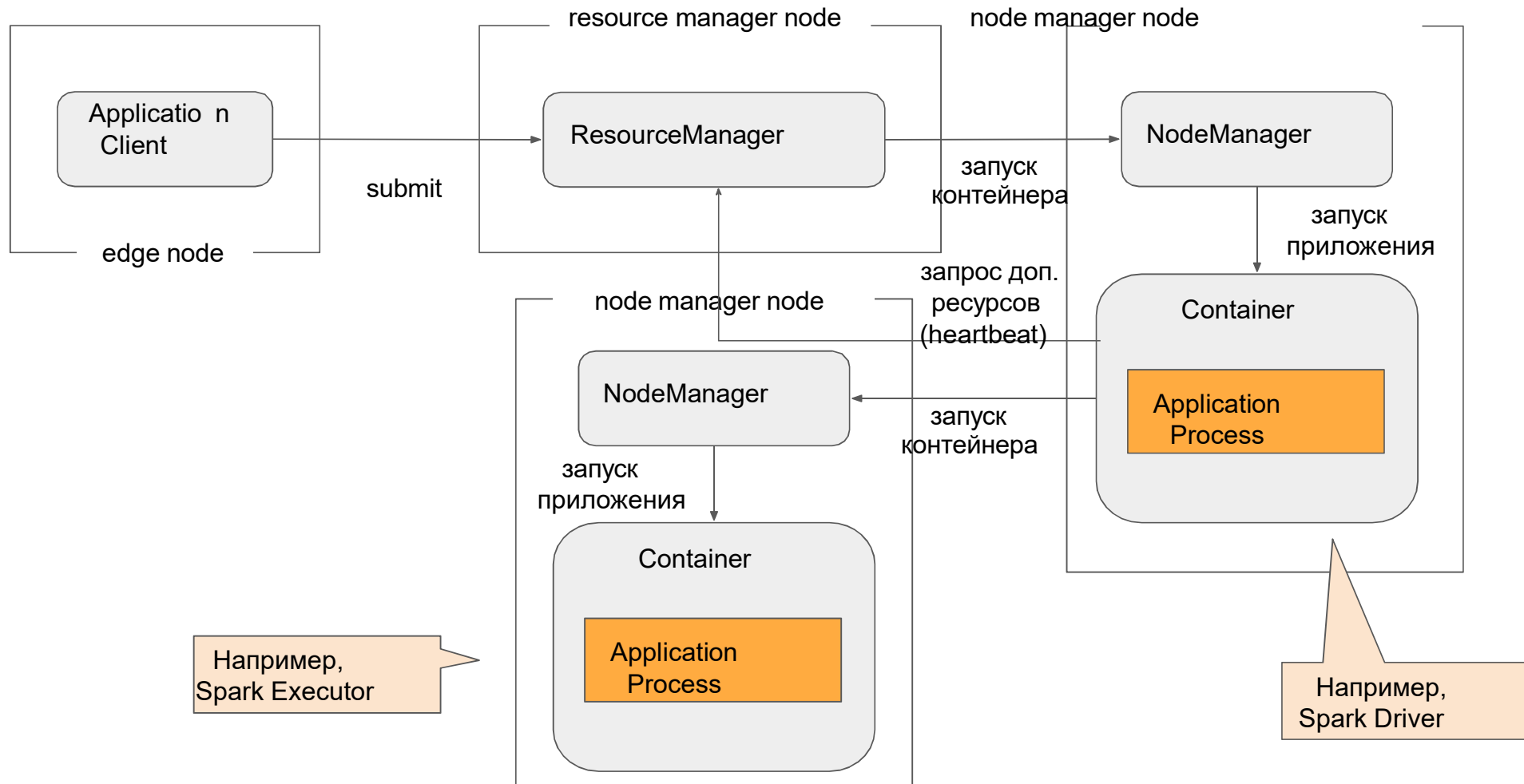
Управление ресурсами: YARN

YARN (“Yet Another Resource Manager”) -
утилита, которая:

1. Распределяет
вычислительные ресурсы
кластера (память и CPU)
приложениям обработки
данных
2. Определяет порядок
выполнения приложений
на кластере



YARN: процесс распределения ресурсов

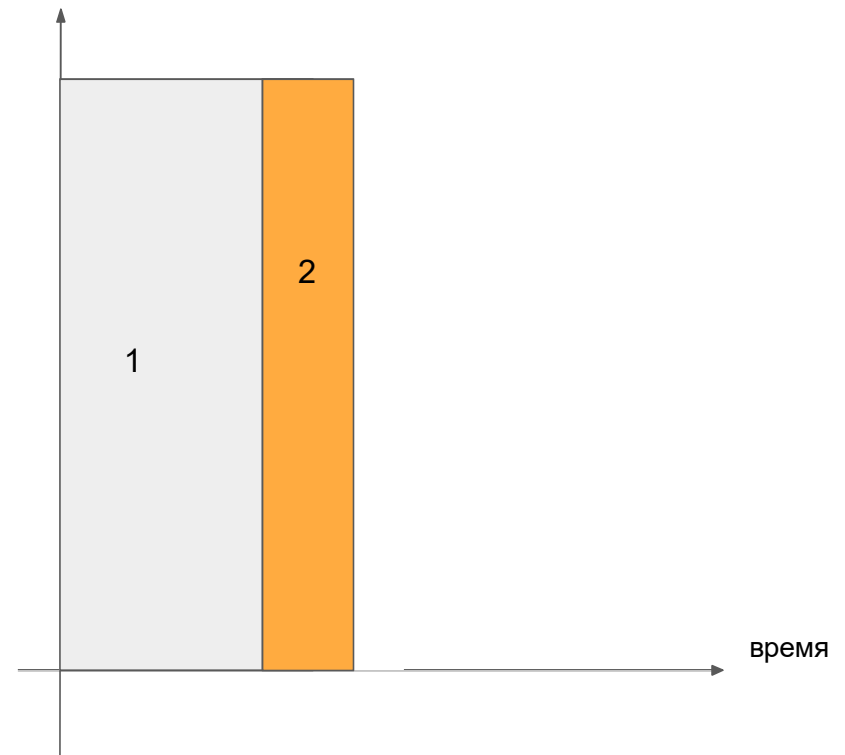


YARN: стратегии распределения ресурсов

1. **FIFO** - “первым пришел, первым обслужен”

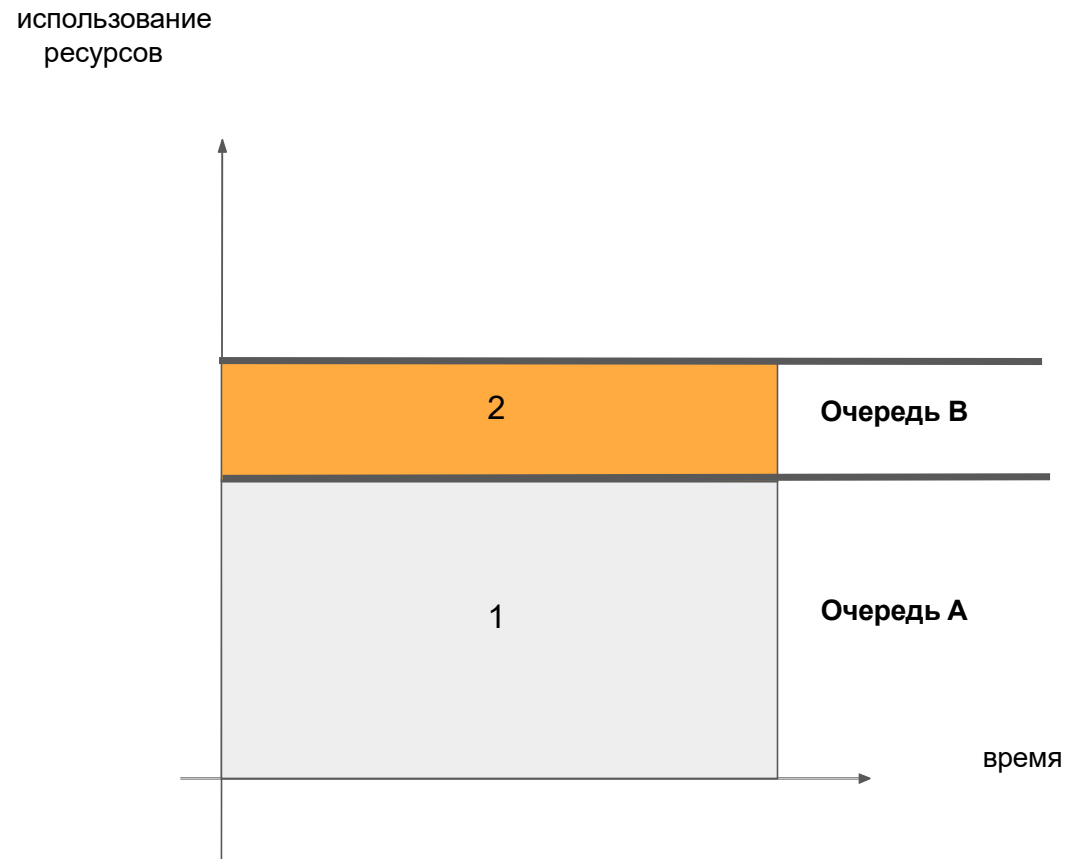
Один процесс занимает все ресурсы в момент времени t . Остальные процессы ждут своей очереди

использование
ресурсов



YARN: стратегии распределения ресурсов

1. **Capacity Scheduler** - ресурсы распределены на пулы (очереди / queue). Один процесс занимает все ресурсы в рамках одной очереди.



YARN: стратегии распределения ресурсов

1. Fair Scheduler - “честное”

распределение.

При появлении второго процесса первый процесс уступает часть ресурсов.



В Spark возможно совместить Fair Scheduler и Capacity Scheduler

использование
ресурсов

