

## 14.1 Финальный тест

Шаг 7 из 25



Каким образом Spark выполняет shuffle (если выполняет) во время операции groupBy и как это влияет на производительность?

Выберите ответ из списка

- Spark реплицирует данные на все вычислительные ноды, что может замедлить вычисления из-за избыточного сетевого трафика
- Spark максимально избегает data shuffling'a путем распределения данных максимально равномерно по партициям, таким образом группировка наиболее эффективна
- Shuffle в Spark перераспределяет данные между партициями, что может быть весьма ресурсоемкой операцией и негативно влиять на производительность
- Spark автоматически избегает shuffle'a (если его явно не затребовать в коде) путем сохранения data locality, который помогает сильно ускорить операции группировки

Отправить

Далее >

