



Текстовая расшифровка видео:

КЭШИРОВАНИЕ

План:

- Промежуточные результаты.

Промежуточные результаты

Если задать вопрос в поисковике о сохранении промежуточных результатов, можно наткнуться на три метода:

- `df.cache;`
- `df.persist;`
- `df.checkpoint.`

Мы говорили о том, что вычисления могут разветвляться. Поскольку все считается «лениво», когда вы каждый раз вызываете расчет, все вычислительное дерево раскрывается с самого начала.

Вы можете сохранить промежуточные результаты с помощью метода «`df.cache`» или «`df.persist`».

Cache для Dataset – это просто алиас на `persist`.

По умолчанию **RDD** кэшируется с **MEMORY_ONLY**, а **DataFrame/Dataset** – с **MEMORY_AND_DISK**.

Кэширование – это **трансформация**, а не действие!

Поясним: у нас есть граф вычисления. Мы считаем до определенного момента, после чего выбираем дальнейший путь. Если на этом моменте мы вызовем `Cache`, то нам все равно хотя бы единожды придется просчитать цепочку. Чтобы `Cache` вызвался нам придется вызвать саму цепочку, то есть вызвать `action`. Сам `Cache` останется в трансформации.

Также существует метод «**`df.unpersist`**», служащий для удаления тех данных, с которыми вы уже



▶ Запустить стенд



Дедлайн 07 июля, 23:59 Мск



Цель чекпоинта – укоротить lineage и избежать потенциальных проблем с переполнением стека.

- **df.checkpoint** сохраняет данные на HDFS.
- **df.localCheckpoint** сохраняет данные в память воркеров.

Если у вас есть длинные отлаженные вычисления, где все отлично работает, то в конце можно добавить либо `df.cache` (если дальше идет разветвление), либо `df.localCheckpoint` (если данные влезают в память). Это позволит проще отлаживать и explain'ить.

- О типах персистентности вы можете ознакомиться, перейдя по ссылке:

<https://spark.apache.org/docs/2.4.8/rdd-programming-guide.html#rdd-persistence>

- С разбором различий вы также можете ознакомиться, перейдя по ссылке:

<https://sparkbyexamples.com/spark/spark-difference-between-cache-and-persist/>

Как вам урок?



Изучил, далее >

Слёрм ©

[+7 \(495\) 248-05-80](tel:+7(495)248-05-80)

[Лицензия №ДЛ-1368 от 22.08.2019](#)

[Политика конфиденциальности](#)

[Публичная оферта](#)