

## RDD的容错

问过的一些公司：端点数据(2021.07)

参考答案：

### 1、Lineage机制（简称血统机制）

RDD的Lineage机制是将粗颗粒度相同的操作（map flatmap filter之类的）记录下来，万一RDD信息丢失可通过Lineage获取信息从新获得原来丢失的RDD

缺点：lineage如果过长 影响性能，开销太大

### 2、checkpoint机制

将RDD的结果存入HDFS，以文件的形式存在

### 3、cache机制

将结果存入内存

### 4、persist机制

将结果存入磁盘

**checkpoint 和 cache 和 persist的异同**

**相同点：**都要在action算子之后触发

**不同点：**

checkpoint：

- 是存在HDFS上的，更加安全可靠，并且checkpoint会改变依赖关系，并且之后会触发一个job操作

persist和cache

- 不改变依赖关系，运行完成后缓存数据自动消失

欢迎加入知识星球，获取《大数据面试题 V4.0》以及更多大数据开发内容



知识星球

长按扫码领取优惠 ▶

