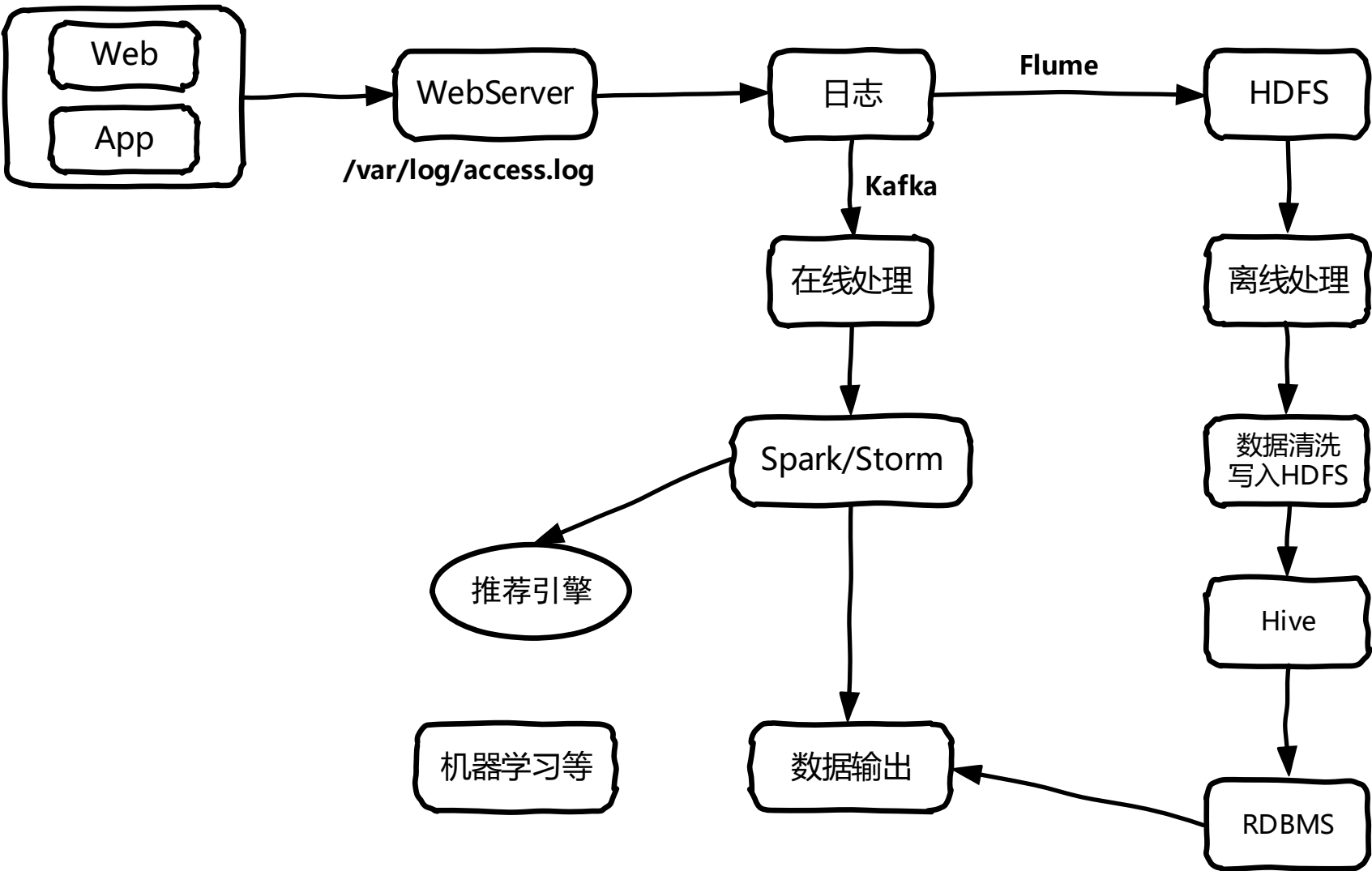
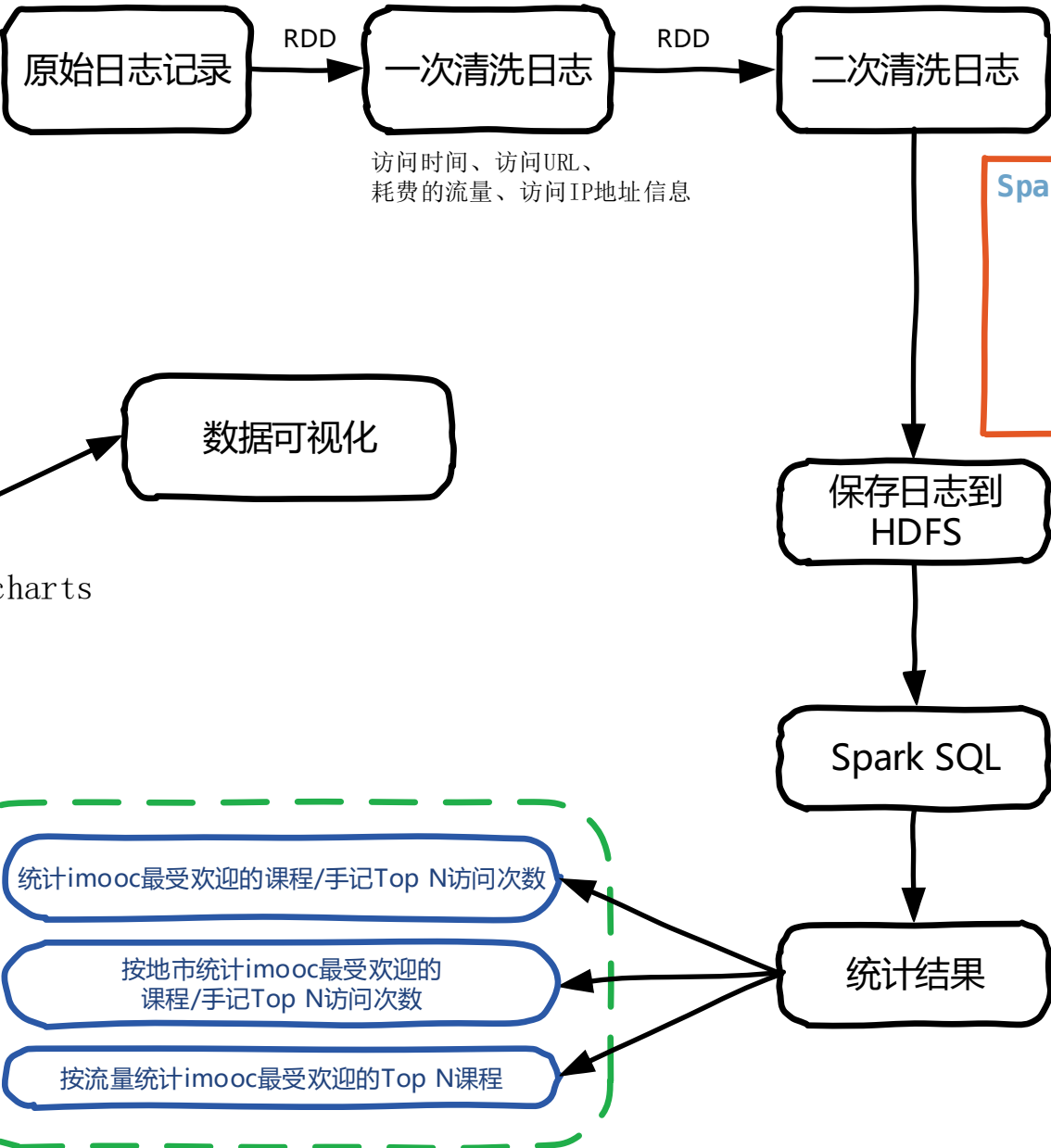


# 离线数据处理架构



## 以慕课网日志分析为例 进入大数据 Spark SQL 的世界

183.162.52.7 - - [10/Nov/2016:00:01:02 +0800] "POST /api3/userdynamic HTTP/1.1" 200 19501 "www.imooc.com" "-" cid=0  
&timestamp=1478707261847&uid=2871142&touid=2871142&page=1&secret=a6e8e14701ffe9f6063934780d9e2e6d&token=3837a5bf27  
ea718fe18bda6c53fbbc14 "mukewang/5.0.0 (Android 5.1.1; Xiaomi Redmi 3 Build/LMY47V), Network 2G/3G"  
10.100.136.6  
5:80 200 0.195 0.195



URL、cmsType(video/article)、cmsId(编号)、流量、ip、城市信息、访问时间、天

SparkStatCleanJob类

day=20170601  
day=20170602  
day=20170603  
day=20170604

TopNStatJob类

利用Spark SQL  
从parquet中读取所有的数据

可以通过DataFrame或者SQL的方式  
进行数据的统计

保存到MySQL中  
StatDao类

