



链接: <https://www.zhihu.com/question/26568496/answer/224439650>

a. 蓝色部分:是Hadoop生态系统组件, 黄色部分是Spark生态组件, 虽然他们是两种不同的大数据处理框架, 但它们不是互斥的, Spark与hadoop 中的MapReduce是一种相互共生的关系。Hadoop提供了Spark许多没有的功能, 比如分布式文件系统, 而Spark 提供了实时内存计算, 速度非常快。有一点大家要注意, Spark并不是一定要依附于Hadoop才能生存, 除了Hadoop的HDFS, 还可以基于其他的云平台, 当然啦, 大家一致认为Spark与Hadoop配合默契最好摆了。

b. 技术趋势: Spark在崛起, hadoop和Storm中的一些组件在消退。大家在学习使用相关技术的时候, 记得与时俱进掌握好新的趋势、新的替代技术, 以保持自己的职业竞争力。

1. HSQL未来可能会被Spark SQL替代, 现在很多企业都是HIVE SQL和Spark SQL两种工具共存, 当Spark SQL逐步成熟的时候, 就有可能替换HSQL;

2. MapReduce也有可能被Spark 替换，趋势是这样，但目前Spark还不够成熟稳定，还有比较长的路要走；

3. Hadoop中的算法库Mahout正被Spark中的算法库MLib所替代，为了不落后，大家注意去学习MLib算法库；

4. Storm会被Spark Streaming替换吗？在这里，Storm虽然不是hadoop生态中的一员，但我仍然想把它放在一起做过比较。由于Spark和hadoop天衣无缝的结合，Spark在逐步的走向成熟和稳定，其生态组件也在逐步的完善，是冉冉升起的新星，我相信Storm会逐步被挤压而走向衰退。