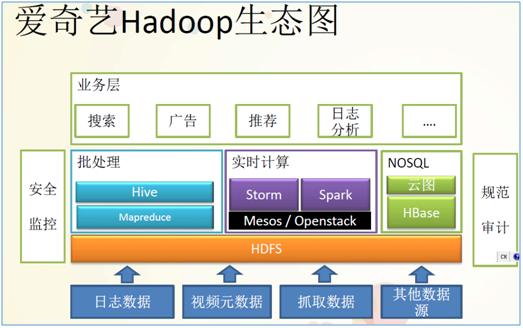
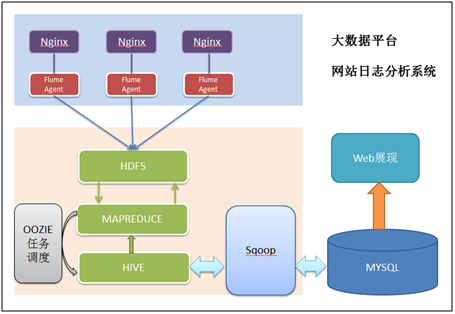
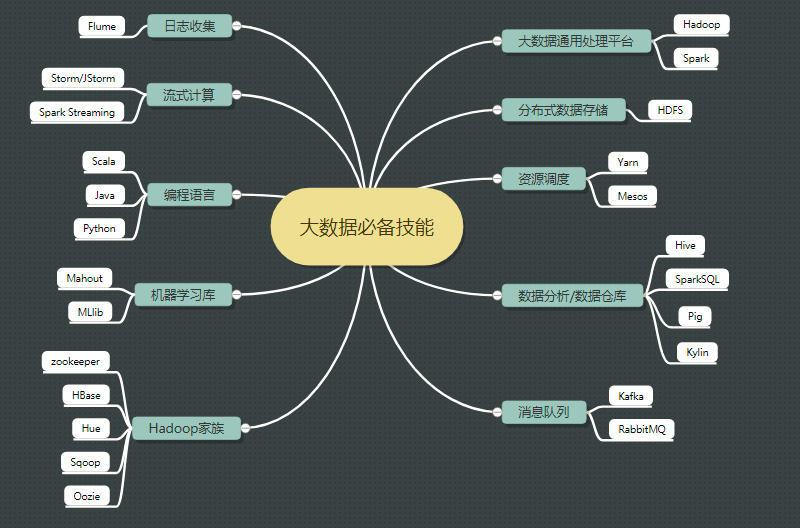
# 、大数据介绍

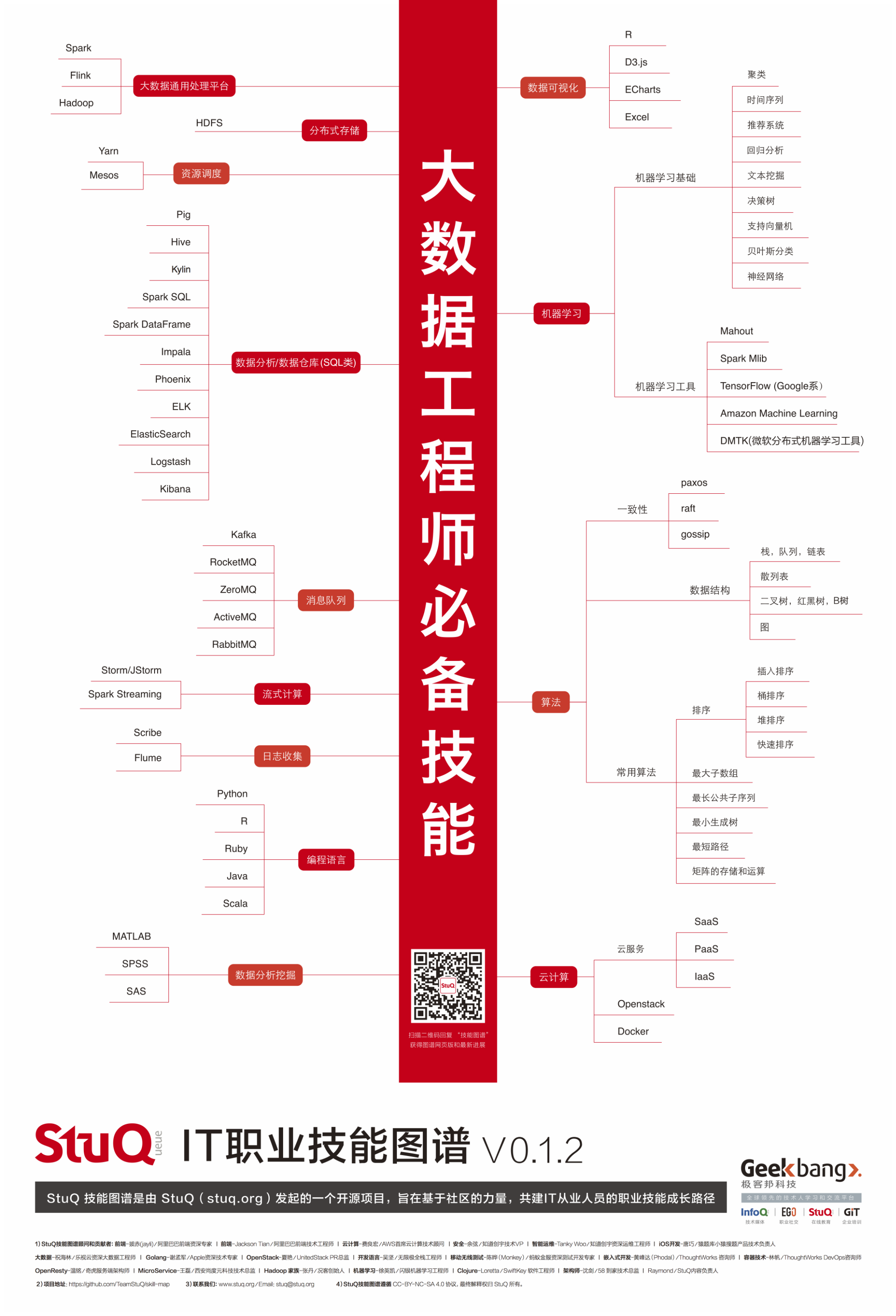
1. 精准营销 – 广告 推荐
2. 商业和社会信用 –蚂蚁信用、蚂蚁花呗

# 、相关技术









文件存储：**Hadoop HDFS** 、Tachyon、 KFS

离线计算：**Hadoop MapReduce**、**Spark**

流式实时计算：**Storm**、Spark Steaming、S4、Heron

K-V、Nosql数据库：**HBase**、Redis、MongoDB

资源管理：**Hadoop YARN**、Mesos

日志收集：**Flume**、Scribe、Logstash、Kibana

消息系统：**Kafka**、StormMQ、ZeroMQ、RabbitMQ

查询分析：**Hive**、Pig、Impala、Presto、Phoenix、SprakSQL、Drill、Flink、Kylin、Druid

分布式协调服务：**Zookeeper**

集群管理及监控：Ambari、Ganglia、Nagios、Cloudera Manager

数据挖掘、机器学习：Mahout、Spark MLLib

数据同步：**Sqoop**

任务调度：Oozie

1. Java高级(虚拟机、并发) –30小时

2. Linux 基本操作 –20小时

3. Hadoop(HDFS+MapReduce+Yarn ) -80小时

4. HBase(JavaAPI操作+Phoenix ) –20小时

5. Hive(Hql基本操作和原理理解) –20小时

6. Kafka

7. Storm/JStorm

8. Scala --20小时

9. Python –30小时

10. Spark (Core+sparksql+Spark streaming ) –60小时

11. 辅助小工具(Sqoop/Flume/Oozie/Hue等)

# 、学习途径

学习资料共享地址：<https://github.com/elfkingw/bigdatasource>

## Hadoop集群（HDFS+MapReduce+Yarn ） 参考学习时间 80h

Hadoop相关视频：<http://edu.51cto.com/center/course/lesson/index?id=183506>

官方文档： <http://hadoop.apache.org/>

中文文档：<http://hadoop.apache.org/docs/r1.0.4/cn/index.html>

书籍：《Hadoop权威指南》

<https://pan.baidu.com/s/1-d1obM93ZtyxCTdGqbU80Q>

# 、学习计划

## 了解相关技术和趋势，掌握学习途径，按阶段制定学习计划

## 第一阶段计划：掌握Hadoop集群（HDFS+MapReduce+Yarn ）部署 12月底

目标：

1. 了解HDFS、MapReduce、Yarn原理
2. 熟练搭建Hadoop集群、能熟练解决集群相关配置问题
3. 搭建简单Demo做MapReduce开发