**Hadoop，HBase，Storm，Spark到底是什么？**

**下面都是Hadoop的子模块**

[**Hadoop**](http://lib.csdn.net/base/hadoop)=HDFS+Hive+Pig+...

**HDFS**: 存储系统  
**MapReduce**：计算系统  
**Hive**：提供给SQL开发人员（通过HiveQL）的MapReduce，基于Hadoop的数据仓库框架  
**Pig**：基于Hadoop的语言开发的  
**HBase**:NoSQL[**数据库**](http://lib.csdn.net/base/mysql)  
**Flume**：一个收集处理Hadoop数据的框架  
**Oozie**：一个让用户以多种语言（如MapReduce，Pig和Hive）定义一系列作业的工作流处理系统  
**Ambari**：一个基于web的部署/管理/监控Hadoop集群的工具集  
**Avro**：允许编码Hadoop文件的schema的一种数据序列化系统  
**Mahout**：一个数据挖掘库，它包含了最流行的一些数据挖据[**算法**](http://lib.csdn.net/base/datastructure)，并且以MapReduce模型来实现他们  
**Sqoop**：一个从非Hadoop数据存储（如关系数据库和数据仓库）进来的移动数据到Hadoop中的连接工具  
**HCatalog**：一个中心化的元数据管理以及Apache Hadoop共享服务，它允许在Hadoop集群中的所有数据的统一视图，并允许不同的工具，包括Pig和Hive，处理任何数据元素，而无需知道身体在集群中的数据存储。

**BigTop**：为了创造一个更正式的程序或框架Hadoop的子项目及相关组件的目标提高Hadoop的平台，作为一个整体的包装和互操作性测试。

**Apache  Storm**：一个分布式实时计算系统，Storm是一个任务并行连续计算引擎。 Storm本身并不典型在Hadoop集群上运行，它使用Apache ZooKeeper的和自己的主/从工作进程，协调拓扑，主机和工作者状态，保证信息的语义。无论如何， Storm必定还是可以从HDFS文件消费或者从文件写入到HDFS。

[**Apache Spark**](http://lib.csdn.net/base/spark)：一种快速，通用引擎用于大规模数据处理，[**Spark**](http://lib.csdn.net/base/spark)是一个数据并行通用批量处理引擎。工作流中在一个类似的和怀旧风格的MapReduce中定义，但是，比传统Hadoop MapReduce的更能干。Apache Spark有其流API项目，该项目通过短间隔批次允许连续处理。Apache Spark本身并不需要Hadoop操作。但是，它的数据并行模式，需要稳定的数据优化使用共享文件系统。该稳定源的范围可以从S3，NFS或更典型地，HDFS。执行Spark应用程序并不需要Hadoop YARN。Spark有自己独立的主/服务器进程。然而，这是共同的运行使用YARN容器Spark的应用程序。此外，Spark还可以在Mesos集群上运行

|  |
| --- |
| Hadoop视频[共44集适合入门 MP4格式]  <https://pan.baidu.com/s/1eROuzvW#path=%252F>  第三阶段 spark  <https://pan.baidu.com/s/1hrH4INy#path=%252F%25E4%25BA%2591%25E8%25AE%25A1%25E7%25AE%2597%252F%25E5%25A4%25A7%25E6%2595%25B0%25E6%258D%25AE%252FHadoop%252F%25E7%25AC%25AC%25E4%25B8%2589%25E9%2598%25B6%25E6%25AE%25B5%2520spark>  淘宝买的hadoop  <https://pan.baidu.com/s/1kVaptiv#path=%252F%25E6%25B7%2598%25E5%25AE%259D%25E4%25B9%25B0%25E7%259A%2584hadoop>  Scala语言入门  <https://pan.baidu.com/s/1o83v9VW>  小象《Hadoop 2.X大数据平台V3》(价值499)  <https://pan.baidu.com/s/1sl4BzL7>  Zookeeper  <https://pan.baidu.com/s/1kTpLLTt>  小象学院《HBase进阶》价值699  <https://pan.baidu.com/s/1gdfK0IR>  小象学院《Hadoop进阶》价值699  <https://pan.baidu.com/s/1mgCKpdu>  玩转大数据：深入浅出大数据挖掘技术培训教程  <https://pan.baidu.com/share/link?shareid=3387739075&uk=506111260#path=%252FHadoop%252Fbigdata>  hadoop高级  <https://pan.baidu.com/share/link?shareid=3387739075&uk=506111260#path=%252FHadoop%252Fhadoop%25E9%25AB%2598%25E7%25BA%25A7>  深入浅出Hive企业级架构优化、Hive Sql优化、压缩和分布式缓存  <https://pan.baidu.com/share/link?shareid=3387739075&uk=506111260#path=%252FHadoop%252Fhive>  Zookeeper入门到精通  <https://pan.baidu.com/share/link?shareid=3387739075&uk=506111260#path=%252FHadoop%252Fzookeeper>  最新云计算(hadoop)  <https://pan.baidu.com/s/1eQgw4Jg#path=%252F%25E6%259C%2580%25E6%2596%25B0%25E4%25BA%2591%25E8%25AE%25A1%25E7%25AE%2597(hadoop)>  800份营销推广活动策划方案.rar  <https://pan.baidu.com/s/1mio8nhQ>  大数据Spark案例100经典案例V2.rar  <https://pan.baidu.com/s/13B2gA>  大数据前置课程—Scala  <https://pan.baidu.com/s/1jGrPJGM#path=%252F>  hadoop  <https://pan.baidu.com/share/link?shareid=1230961459&uk=3591204871#path=%252F>  Hadoop视频教程  <https://pan.baidu.com/share/link?shareid=3289532506&uk=3456295149#path=%252FHadoop%25E8%25A7%2586%25E9%25A2%2591%25E6%2595%2599%25E7%25A8%258B> |