

## 6-hadoop体系技术

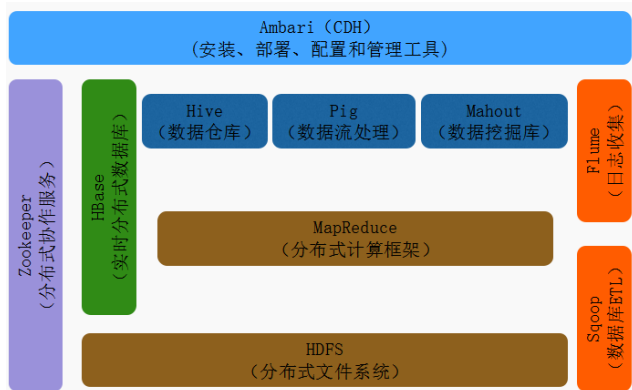
-----成都尚学堂-mr-zeng-----

### Hadoop是什么

是一个由Apache基金会所开发的分布式系统基础架构。

### Hadoop体系的技术

**核心：**hdfs（存储），mapreduce（计算）



### hdfs

分布式存储框架（**善于存储大文件**）

### mapreduce

分布式计算框架

### yarn

分布式计算所需要的各节点资源管理框架,任务调度与资源管理

### hive

类似于sql的方式操作hadoop的mapreduce进行计算，底层会转换为mapreduce代码

### pig

hadoop提供的操作mapreduce的新的编程语言，底层会转换为mapreduce代码

### hbase

基础与hdfs的，分布式面向列的数据库（**善于存储小文件**）

### sqoop

提供了hdfs和关系型数据库之间进行数据导入导出

### flume

提供了海量日志采集，聚合，传输，分布式存储服务。

### mahout

数据挖掘工具，分布式机器学习的算法集合，把以前单节点的算法实现转换为mapreduce的实现-运行在集群上-提供数据处理量和运算的性能。

### cdh与ambari

提供了对hadoop体系各技术进行统一配置，服务，管理的工具。

### zookeeper

分布式的协调服务。

功能：

- a) 提供的小型分布式文件系统（**内存中-访问速度快，存储量不大**）。
- b) 提供了对文件系统里各数据操作的监听事件。

通过上述功能通常的应用：

- a) 服务器单点的高可用性。（**最重要的功能**）
- b) 分布式配置文件统一管理。

...

...