# 阐述HDFS和HBase在Hadoop生态中的功能和联系，以及未来技术展望

## 功能：

**HDFS的功能**为数据的存储、管理和出错处理。设计的目的是用于可靠地存储大规模的数据集，并提高用户访问数据的效率。

它主要用于：

1.适合大文件存储和处理：可处理的文件规模可达到百MB乃至数百TB，目前应用已到PB级。

2..集群规模可动态扩展::存储节点可在运行状态下加入到集群中，集群仍然可以正常地工作。

3.能有效保证数据一致性:基于“一次写入，多次读取”设计，简化处理文件访问方式，当一个文件创建、写入并关闭后就不能再修改。

4:数据的吞吐量大跨平台移植性好:采用数据流式读写的方式，用以增加数据的吞吐量。具有很好的跨平台移植性，源代码开放。

**HBase功能：可以**用来存储半结构化和非结构化的数据，能提供高可靠性、高性能、面向列、可伸缩、实时读写的分布式数据库。

HBase 它底层的文件系统是通过 HDFS 来支持，具有HDFS的功能。此外HBase 在 HDFS 之上提供了:

1：高并发实时随机写，通过 LSM(内存+顺序写磁盘)的方式提供了 HDFS 所不拥有的实时随机写及修改功能

2：高并发实时点读及扫描了解一下 LSM 算法，在文件系统之上有数据库，在业务层面，HBase 完全可以独立于 HDFS 来理解

## 联系

HBase是Hadoop的子项目，建立在HDFS之上，原型为Google的分布式存储系统BigTable。HBase的数据通常存储在HDFS上。HDFS为HBase提供了高可靠性的底层存储支持。HBase依赖于HDFS。

同时从上面的介绍中，我们可以看到HBASE底层的文件系统是通过 HDFS 来支持，具有HDFS的功能，两者相互支持，使得HBASE能够完整更加复杂的功能任务和业务。

## 未来技术的展望

10年以后的Hadoop应该只是一个生态的“代名词”了，这很容易理解，就像十年前的人们想象不到如今手机电脑的普及，科技发展的如此迅速一样，HDFS和HBase在Hadoop生态中仍然扮演着非常重要的角色。未来，HDFS将继续发展以满足更高的可靠性、可扩展性和性能需求。同时我推测存储层不只是HDFS、HBase和等现有的存储架构，对于处理组件更会像app store里的应用一样多，任何第三方都可以根据Hadoop的数据访问等开发自己的组件，用户在市场中根据自己数据的使用特性和计算需求选择相应的组件运用。

最后，SQL和NOSQL可能会合并，NewSQL和Hadoop技术相互借鉴而最终走向统一，Hadoop市场和数据仓库市场会合并，形成规模更大更具有影响力和市场的产品。Hadoop与云平台集成，融合一些前沿技术统一资源调度管理，提供强力能力，企业数据分析中心合并为单一架构。企业大数据产品场景化。以后直接提供产品和技术的公司趋于成熟并目转向服务。越来越多的新公司提供的是行业化、场景化的解决方案，如一些特定的需求。