

选择题 181250035 高金比象

ABBAAD CADDB ABBAAC

判断题

X X X X V V V X

简答题

1. Volume (大体量): 即可从数百TB到数千PB甚至EB  
规模

Variety (多样性): 即大数据包括各种种类和形态  
的数据

Velocity (时效性): 即很多大数据需要在一定的  
时间限度下得到及时处理

Veracity (准确性): 即处理的结果要保证一定的  
准确性

Value (大价值): 即大数据包含很多潜在的价值, 好  
数据分析挖掘和利用带来巨大的商业价值

2. name node  $\rightarrow$  data node  $\rightarrow$  secondary name node  $\rightarrow$   
resource manager  $\rightarrow$  node manager

3. (1) 列式存储 (2) 大数据是稀疏的多维映射表

(3) 读写的一致性 (4) 提供很高的数据读写速度

(5) 良好的线性可扩展性 (6) 提供海量数据

(7) 数据会自动分片 (8) 对于数据故障, hbase是有自动

的负载均衡和恢复能力 (9) 提供了方便的与HDFS和  
MAP REDUCE集成能力,

论述题:

1. Map Reduce并行计算框架。MapReduce并行计算框架是个并行  
化程序执行系统。它提供了一个包含Map和Reduce两阶段  
的并行处理模型和过程, 提供一个并行处理模型和接口,  
让程序员可以方便快速地编写出大数据并行处理程序。



2. 分布式文件系统HDFS。HDFS (Hadoop Distributed File System) 是一个类似于Google GFS 的开源的分布式文件系统。它提供了一个可扩展、高可靠、高可用的大规模数据分布式存储管理系统，基本上分布在各节点数据存储空间上的本地Linux系统的文件系统，在应用程序提供了个逻辑上成为整体的大规模数据存储系统。

3. 分布式数据库管理系统HBase  
为了克服HDFS难以管理结构化、半结构化海量数据的缺点，Hadoop 提供了一个大规模分布式数据库管理和查询系统HBase。HBase是一个建立在HDFS之上的分布式数据库，它是一个分布式可扩展的NoSQL数据库，提供了对结构化、半结构化甚至非结构化大数据实时读写和随机访问能力。

4. 公共服务框架Common。Common是一套为整个Hadoop系统提供底层支撑服务和常用工具类库和API编程接口。这些底层服务包括Hadoop抽象文件系统File System、远程过程调用RPC、系统配置工具Configuration以及序列化机制。

5. 分布式协调服务框架Zookeeper。Zookeeper是<sup>分布式</sup>一个协调服务框架，主要用于解决分布式环境中的一致性问题。Zookeeper主要用于提供分布式应用中经常需要的系统可靠协调、数据状态同步、统一命名服务、分布式应用配置项管理等功能。

6. 分布式数据仓库处理工具Hive。Hive是一个建立在Hadoop之上的数据仓库，用于管理存储在HDFS或HBase中的结构化/半结构化数据。它最早由Facebook开发并用于处理分析大量的用户日志数据，2008年Facebook将其贡献给Apache成为Hadoop项目。