



## 前言：建议 Word 原版，使用“查找”功能查题

# 大数据技术原理与应用（林子雨）

作者：解忧书店 JieYouBookshop

## 第 1 章 大数据概述

### 1 单选(2分)

第三次信息化浪潮的标志是：

- A. 个人电脑的普及
- B. 云计算、大数据、物联网技术的普及
- C. 虚拟现实技术的普及
- D. 互联网的普及

正确答案：B 你选对了

### 2 单选(2分)

就数据的量级而言，1PB 数据是多少 TB？

- A. 2048
- B. 1000
- C. 512
- D. 1024

正确答案：D 你选对了

### 3 单选(2分)

以下关于云计算、大数据和物联网之间的关系，论述错误的是：

- A. 云计算侧重于数据分析
- B. 物联网可以借助于云计算实现海量数据的存储
- C. 物联网可以借助于大数据实现海量数据的分析
- D. 云计算、大数据和物联网三者紧密相关，相辅相成

正确答案：A 你选对了

### 4 单选(2分)

以下哪个不是大数据时代新兴的技术：

- A. Spark
- B. Hadoop
- C. HBase
- D. MySQL

正确答案：D 你选对了



5 单选 (2 分)

每种大数据产品都有特定的应用场景，以下哪个产品是用于批处理的：

A.MapReduce

B.Dremel

C.Storm

D.Pregel

正确答案：A 你选对了

6 单选 (2 分)

每种大数据产品都有特定的应用场景，以下哪个产品是用于流计算的：

A.GraphX

B.S4

C.Impala

D.Hive

正确答案：B 你选对了

7 单选 (2 分)

每种大数据产品都有特定的应用场景，以下哪个产品是用于图计算的：

A.Pregel

B.Storm

C.Cassandra

D.Flume

正确答案：A 你选对了

8 单选 (2 分)

每种大数据产品都有特定的应用场景，以下哪个产品是用于查询分析计算的：

A.HDFS

B.S4

C.Dremel

D.MapReduce

正确答案：C 你选对了

9 多选 (3 分)

数据产生方式大致经历了三个阶段，包括：

A.运营式系统阶段

B.感知式系统阶段

C.移动互联网数据阶段



D.用户原创内容阶段

正确答案：ABD你选对了

10 多选(3分)

大数据发展的三个阶段是：

- A.低谷期
- B.成熟期
- C.大规模应用期
- D.萌芽期

正确答案：BCD你选对了

11 多选(3分)

大数据的特性包括：

- A.价值密度低
- B.处理速度快
- C.数据类型繁多
- D.数据量大

正确答案：ABCD你选对了

12 多选(3分)

图领奖获得者、著名数据库专家 Jim Gray 博士认为，人类自古以来在科学研究上先后经历了哪几种范式：

- A.计算科学
- B.数据密集型科学
- C.实验科学
- D.理论科学

正确答案：ABCD你选对了

13 多选(3分)

大数据带来思维方式的三个转变是：

- A.效率而非精确
- B.相关而非因果
- C.精确而非全面
- D.全样而非抽样

正确答案：ABD你选对了

14 多选(3分)

大数据主要有哪几种计算模式：



- A.流计算
- B.图计算
- C.查询分析计算
- D.批处理计算

正确答案：ABCD你选对了

15 多选 (3 分)

云计算的典型服务模式包括三种：

- A.SaaS
- B.IaaS
- C.MaaS
- D.PaaS

正确答案：ABD你选对了

## 第 2 章 大数据处理架构 Hadoop

1 单选 (2 分)

启动 hadoop 所有进程的命令是：

- A.start-dfs.sh
- B.start-all.sh
- C.start-hadoop.sh
- D.start-hdfs.sh

正确答案：B你选对了

2 单选 (2 分)

以下对 Hadoop 的说法错误的是：

- A.Hadoop 是基于 Java 语言开发的，只支持 Java 语言编程
- B.Hadoop2.0 增加了 NameNode HA 和 Wire-compatibility 两个重大特性
- C.Hadoop MapReduce 是针对谷歌 MapReduce 的开源实现，通常用于大规模数据集的并行计算
- D.Hadoop 的核心是 HDFS 和 MapReduce

正确答案：A你选对了

3 单选 (2 分)

以下哪个不是 Hadoop 的特性：

- A.成本高
- B.支持多种编程语言



C.高容错性

D.高可靠性

正确答案：A你选对了

4 单选 (2 分)

以下名词解释不正确的是：

A.Zookeeper：针对谷歌 Chubby 的一个开源实现，是高效可靠的协同工作系统

B.HBase：提供高可靠性、高性能、分布式的行式数据库，是谷歌 BigTable 的开源实现

C.Hive：一个基于 Hadoop 的数据仓库工具，用于对 Hadoop 文件中的数据集市进行数据整理、特殊查询和分析存储

D.HDFS 分布式文件系统，是 Hadoop 项目的两大核心之一，是谷歌 GFS 的开源实现

正确答案：B你选对了

5 多选 (3 分)

以下哪些组件是 Hadoop 的生态系统的组件：

A.HBase

B.Oracle

C.HDFS

D.MapReduce

正确答案：ACD你选对了

6 多选 (3 分)

以下哪个命令可以用来操作 HDFS 文件：

A.hadoop fs

B.hadoop dfs

C.hdfs fs

D.hdfs dfs

正确答案：ABD你选对了

### 第 3 章 分布式文件系统 HDFS

1 单选 (2 分)

HDFS 的命名空间不包含：

A.字节

B.文件

C.块



## D.目录

正确答案：A你选对了

### 2 单选 (2 分)

对 HDFS 通信协议的理解错误的是：

- A. 客户端与数据节点的交互是通过 RPC (Remote Procedure Call ) 来实现的
- B. 客户端通过一个可配置的端口向名称节点主动发起 TCP 连接, 并使用客户端协议与名称节点进行交互
- C. 名称节点和数据节点之间则使用数据节点协议进行交互
- D. HDFS 通信协议都是构建在 IoT 协议基础之上的

正确答案：D你选对了

### 3 单选 (2 分)

采用多副本冗余存储的优势不包含：

- A. 保证数据可靠性
- B. 容易检查数据错误
- C. 加快数据传输速度
- D. 节约存储空间

正确答案：D你选对了

### 4 单选 (2 分)

假设已经配置好环境变量, 启动 Hadoop 和关闭 Hadoop 的命令分别是：

- A. start-dfs.sh , stop-hdfs.sh
- B. start-hdfs.sh , stop-hdfs.sh
- C. start-dfs.sh , stop-dfs.sh
- D. start-hdfs.sh , stop-dfs.sh

正确答案：C你选对了

### 5 单选 (2 分)

分布式文件系统 HDFS 采用了主从结构模型, 由计算机集群中的多个节点构成的, 这些节点分为两类, 一类存储元数据叫 , 另一类存储具体数据叫 :

- A. 名称节点, 主节点
- B. 从节点, 主节点
- C. 名称节点, 数据节点
- D. 数据节点, 名称节点

正确答案：C你选对了

### 6 单选 (2 分)





下面关于分布式文件系统 HDFS 的描述正确的是：

- A. 分布式文件系统 HDFS 是 Google Bigtable 的一种开源实现
- B. 分布式文件系统 HDFS 是谷歌分布式文件系统 GFS (Google File System) 的一种开源实现
- C. 分布式文件系统 HDFS 比较适合存储大量零碎的小文件
- D. 分布式文件系统 HDFS 是一种关系型数据库

正确答案：B 你选对了

7 多选 (3 分)

以下对名称节点理解正确的是：

- A. 名称节点作为中心服务器，负责管理文件系统的命名空间及客户端对文件的访问
- B. 名称节点用来负责具体用户数据的存储
- C. 名称节点通常用来保存元数据
- D. 名称节点的数据保存在内存中

正确答案：ACD 你选对了

8 多选 (3 分)

以下对数据节点理解正确的是：

- A. 数据节点通常只有一个
- B. 数据节点用来存储具体的文件内容
- C. 数据节点的数据保存在磁盘中
- D. 数据节点在名称节点的统一调度下进行数据块的创建、删除和复制等操作

正确答案：BCD 你选对了

9 多选 (3 分)

HDFS 只设置唯一一个名称节点带来的局限性包括：

- A. 集群的可用性
- B. 性能的瓶颈
- C. 命名空间的限制
- D. 隔离问题

正确答案：ABCD 你选对了

10 多选 (3 分)

以下 HDFS 相关的 shell 命令不正确的是：

- A. `hadoop dfs mkdir <path>` : 创建 <path> 指定的文件夹
- B. `hdfs dfs -rm <path>` : 删除路径 <path> 指定的文件
- C. `hadoop fs -copyFromLocal <path1> <path2>` : 将路径 <path2> 指定的文件或文件夹复制到路径 <path1> 指定的文件夹中



D.hadoop fs -ls <path> : 显示<path>指定的文件的详细信息

正确答案: AC你选对了

## 第 4 章 分布式数据库 HBase

1 单选 (2 分)

HBase 是一种 数据库

- A.行式数据库
- B.关系数据库
- C.文档数据库
- D.列式数据库

正确答案: D你选对了

2 单选 (2 分)

下列对 HBase 数据模型描述错误的是:

- A.每个 HBase 表都由若干行组成，每个行由行键 (row key) 来标识
- B.HBase 是一个稀疏、多维度、排序的映射表，这张表的索引是行键、列族、列限定符和时间戳
- C.HBase 中执行更新操作时，会删除数据旧的版本，并生成一个新的版本
- D.HBase 列族支持动态扩展，可以很轻松地添加一个列族或列

正确答案: C你选对了

3 单选 (2 分)

下列说法正确的是:

- A.如果不启动 Hadoop，则 HBase 完全无法使用
- B.HBase 的实现包括的主要功能组件是库函数，一个 Master 主服务器和一个 Region 服务器
- C.如果通过 HBase Shell 插入表数据，可以插入一行数据或一个单元格数据
- D.Zookeeper 是一个集群管理工具，常用于分布式计算，提供配置维护、域名服务、分布式同步等

正确答案: D你选对了

4 单选 (2 分)

在 HBase 数据库中，每个 Region 的建议最佳大小是:

- A.2GB-4GB
- B.100MB-200MB
- C.500MB-1000MB
- D.1GB-2GB

正确答案: D你选对了





5 单选 (2 分)

HBase 三层结构的顺序是：

- A.Zookeeper 文件，.META表，-ROOT表
- B.-ROOT表，Zookeeper 文件，.META表
- C.Zookeeper 文件，-ROOT表，.META表
- D.META表，Zookeeper 文件，-ROOT表

正确答案：C你选对了

6 单选 (2 分)

客户端是通过 级寻址来定位 Region:

- A.三
- B.二
- C.一
- D.四

正确答案：A你选对了

7 单选 (2 分)

关于 HBase Shell 命令解释错误的是：

- A.create : 创建表
- B.put : 向表、行、列指定的单元格添加数据
- C.list : 显示表的所有数据
- D.get : 通过表名、行、列、时间戳、时间范围和版本号来获得相应单元格的值

正确答案：C你选对了

8 多选 (3 分)

下列对 HBase 的理解正确的是：

- A.HBase 是针对谷歌 BigTable 的开源实现
- B.HBase 是一种关系型数据库，现成功应用于互联网服务领域
- C.HBase 是一个行式分布式数据库，是 Hadoop 生态系统中的一个组件
- D.HBase 多用于存储非结构化和半结构化的松散数据

正确答案：AD你选对了

9 多选 (3 分)

HBase 和传统关系型数据库的区别在于哪些方面：

- A.数据操作
- B.数据索引
- C.数据模型



#### D.存储模式

正确答案：ABCD你选对了

10 多选(3分)

访问 HBase 表中的行，有哪些方式：

- A.通过某列的值区间
- B.全表扫描
- C.通过一个行健的区间来访问
- D.通过单个行健访问

正确答案：BCD你选对了

## 第 5 章 NoSQL数据库

1 单选(2分)

下列关于 NoSQL数据库和关系型数据库的比较，不正确的是：

- A.NoSQL数据库很容易实现数据完整性，关系型数据库很难实现数据完整性
- B.NoSQL数据库缺乏统一的查询语言，而关系型数据库有标准化查询语言
- C.NoSQL数据库的可扩展性比传统的关系型数据库更好
- D.NoSQL数据库具有弱一致性，关系型数据库具有强一致性

正确答案：A你选对了

2 单选(2分)

以下对各类数据库的理解错误的是：

- A.键值数据库的键是一个字符串对象，值可以是任意类型的数据，比如整型和字符型等
- B.文档数据库的数据是松散的，XML和 JSON文档等都可以作为数据存储文档数据库中
- C.图数据库灵活性高，支持复杂的图算法，可用于构建复杂的关系图谱
- D.HBase数据库是列族数据库，可扩展性强，支持事务一致性

正确答案：D你选对了

3 单选(2分)

下列数据库属于文档数据库的是：

- A.MySQL
- B.Redis
- C.MongoDB
- D.HBase

正确答案：C你选对了



4 单选 (2 分)

NoSQL数据库的三大理论基石不包括：

- A.最终一致性
- B.BASE
- C.ACID
- D.CAP

正确答案：C你选对了

5 多选 (3 分)

关于 NoSQL数据库和关系数据库，下列说法正确的是：

- A.NoSQL数据库可以支持超大规模数据存储，具有强大的横向扩展能力
- B.NoSQL数据库和关系数据库各有优缺点，但随着 NoSQL的发展，终将取代关系数据库
- C.大多数 NoSQL数据库很难实现数据完整性
- D.关系数据库有关系代数理论作为基础，NoSQL数据库没有统一的理论基础

正确答案：ACD你选对了

6 多选 (3 分)

NoSQL数据库的类型包括：

- A.键值数据库
- B.列族数据库
- C.文档数据库
- D.图数据库

正确答案：ABCD你选对了

7 多选 (3 分)

CAP是指：

- A.一致性
- B.可用性
- C.持久性
- D.分区容忍性

正确答案：ABD你选对了

8 多选 (3 分)

NoSQL数据库的 BASE特性是指：

- A.软状态
- B.持续性
- C.最终一致性



D.基本可用

正确答案：ACD你选对了

## 第 6 章 云数据库

1 单选 (2 分)

下列 Amazon 的云数据库属于关系数据库的是：

A.Amazon SimpleDB

B.Amazon DynamoDB

C.Amazon RDS

D.Amazon Redshift

正确答案：C你选对了

2 单选 (2 分)

下列关于 UMP 系统的说法不正确的是：

A.Controller 服务器向 UMP 集群提供各种管理服务，实现集群成员管理、元数据存储等功能

B.Agent 服务器部署在运行 MySQL 进程的机器上，用来管理每台物理机上的 MySQL 实例

C.UMP 系统是低成本和高性能的 MySQL 云数据库方案

D.Mnesia 是 UMP 系统的一个组件，是一个分布式数据库管理系统，且不支持事务

正确答案：D你选对了

3 多选 (3 分)

UMP 依赖的开源组件包括

A.LVS

B.ZooKeeper

C.Mnesia

D.RabbitMQ

正确答案：ABCD你选对了

4 多选 (3 分)

在 UMP 系统中，Zookeeper 主要发挥的作用包括：

A.监控所有 MySQL 实例

B.负责集群负载均衡

C.提供分布式锁，选出一个集群的“总管”

D.作为全局的配置服务器

正确答案：ACD你选对了

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/767153052130006025>