**四川工商职业技术学院2020-2021学年第1学期《数据采集系统》期末考试A卷**

**参考答案及评分标准**

**解答教师: 任 亮**

**命题教师：** 任 亮 **使用班级：** **教研室主任(签字)：** **系主任(签字)**：

装 订 线

|  |
| --- |
| **一.单项选择题(每题1分,共16分)**  1. D 2. B 3. A 4. A 5. B 6. D 7. D 8. A 9. C 10. C 11. C 12. D 13. D 14. A 15. A 16. C  **二. 多项选择题（每题3分，共30分）**  1. ABCD 2. AB 3. ABCDE 4. ABCDE 5. ABCD 6. ABCDE 7. ABCD 8. ABCE 9. BDE 10. BCDE  **三. 判断题（每题1分，共10分）**  1. √ 2. √ 3. √ 4.× 5. × 6.× 7. √ 8. √ 9.√ 10. ×  **四.填空题(每空1分,共20分)**  1.HTML, CSS, JavaScript  2.有序列表, 按照顺序, 偏移量  3.外部数据采集，数据管道，网络爬虫，消息中间件  4.特定语境，用户，广告  5.高吞吐，超大文件  6.用户, 组别, 模式  7.leader副本， follower副本  **五.简答题(每题6分，共24分）**  **1.**大数据的基本特征可以用4个V总结，数据量体量巨大（Volume）、数据类型繁多（Variety）、价值（Value）、处理速度快（Velocity）。  **2.**网络爬虫的基础工作流程如下：  1. 获取种子URL  2. 将URL放入待抓取的队列  3. 爬虫解析DNS得到主机IP  4. 根据IP下载网页内容  5. 爬虫解析网页内容，获取价值数据  6. 重复进行种子URL的抓取工作  **3.**Kafka的三个特点是：发布和订阅消息；具备消息的存储和容错能力；即时处理消息  **4.**新的理念之一是“数据驱动”，即经营管理决策可以自下而上地由数据驱动；其二是“数据闭环”，如互联网行业往往能够构造包括数据采集、建模分析、效果评估到反馈修正各个环节在内的完整“数据闭环”，从而能够不断自我升级。 |

四川工商职业技术学院试卷用纸，共 1页，第 1页