附件4

2024年度广西壮族自治区公安机关大数据职位专业科目测试和专业技能测试大纲

为便于报考人员充分了解广西壮族自治区2024年公安机关大数据职位专业科目测试和专业技能测试，特制定本大纲。

一、考试性质和目标

大数据职位专业科目测试和专业技能测试是针对计算机相关学科报考人员设置的专业技术水平测试。考试主要测试报考人员的大数据实操能力素质，包括对云计算、大数据挖掘分析、机器学习等学科相关内容的掌握程度和应用相关知识解决实战问题的能力。

通过考试的人员，表明其已具备从事公安机关大数据职位相应专业岗位工作的能力和水平，用人单位可根据工作需要从通过人员中择优安排从事公安机关大数据系统建设管理、大数据分析研判、信息网络安全防护等专业技术工作。

二、考试方式和时限

**专业科目测试**采用闭卷笔试考试方式，分为客观性试题和主观性试题两部分，考试时限120分钟，满分100分。

**专业技能测试**为计算机实机操作，类似解题模式的CTF竞赛，通过完成下列类似任务来获得分值，如分析指定文件中的一系列数据、在指定的靶场环境（虚拟机、容器）中找到隐藏的“旗帜”（flags）、完成指定任务的操作并反馈结果、编写代码并运行后反馈结果。完成实机操作后，必须按照要求完成电子答卷提交，考官阅卷评分以报考人员提交的电子答卷及附件文件为准。考试时间为180分钟，满分100分。

考试上机环境为Windows 10，虚拟机环境为Kali Linux，集成开发环境（IDE）为Visual Studio、 Eclipse、PyCharm等。

三、考试内容

本大纲规定了每一个知识要点的内容和深度要求，分为“了解”“理解”和“掌握”三类：

了解：是最低深度要求，报考人员需要正确认识该知识要点的基本概念和原理，大纲内未作特别标注的都为了解性要求。

理解（用“★”标注）：是中等深度要求，报考人员需要在正确认识该知识要点的基本概念和原理的基础上，深入理解其内容，并可以进一步地判断、推理，能通过计算机操作完成相关题目。

掌握（用“★★”标注）：是最高深度要求，报考人员需要正确认识该知识要点的概念、原理，并在深入理解的基础上灵活运用、举一反三，并能熟练通过计算机操作完成相关题目。

第一编 法律法规基础知识

1. 大数据相关法律法规
2. 《中华人民共和国网络安全法》
3. 《中华人民共和国数据安全法》
4. 《中华人民共和国个人信息保护法》
5. 《信息安全等级保护安全办法》
6. 《关键信息基础设施安全保护条例》
7. 《贯彻落实网络安全等保制度和关保制度的指导意见》（公网安〔2020〕1960号）

第二编 计算机专业基础知识

1. 计算机组成原理
2. 计算机硬件的发展历程
3. 计算机软件的分类
4. 计算机硬件的基本组成
5. 计算机软件和硬件的关系
6. 计算机的工作过程
7. ★计算机的性能指标
8. ★进位计数制及其相互转换
9. ★字符和字符串
10. 浮点数的表示和运算
11. 存储器的分类
12. ★存储器的性能指标
13. 虚拟存储器
14. 指令的基本格式
15. 常见数据寻址方式
16. CISC和 RISC 的基本概念
17. CPU 的功能和基本结构
18. 总线标准
19. ★常见编码的识别与转换
20. ASCII编码
21. Base64编码
22. shellcode编码
23. Quoted-printable编码
24. XXencode编码
25. UUencode编码
26. URL编码
27. Unicode编码
28. Escape/Unescape编码
29. HTML实体编码
30. 莫尔斯电码（Morse Code）
31. 操作系统
32. 操作系统的基本概念
33. 操作系统的主要功能和提供的服务
34. 操作系统的分类
35. ★进程与线程
36. 处理器的调度
37. ★内存管理，熟悉获取内存数据的方法，可以使用工具将当前主机内存保存为镜像文件。
38. ★文件管理
39. ★设备管理
40. ★★磁盘阵列的主要实现方式
41. ★★Linux基础
42. 常用基础命令
43. 根目录结构
44. 进程管理
45. UID 和 GID
46. 权限设置
47. procfs文件系统
48. 用户管理
49. 引导和启动
50. 基于关键词或者属性条件对文件进行过滤，基于字符串或正则表达式对文件内容进行数据搜索。
51. ★★Windows基础
52. Windows的版本
53. 引导和启动
54. 用户管理
55. 文件系统
56. 服务
57. 通过注册表对文件和历史记录进行搜集和痕迹分析
58. 进程管理
59. 设备管理
60. 计算机网络
61. 计算机网络的组成
62. 计算机网络的分类
63. ★计算机网络体系结构与参考模型
64. 物理层
65. 数据传输方式
66. 传输介质分类
67. 物理层设备
68. 数据链路层
69. 局域网的基本概念与体系结构
70. ★以太网的工作原理
71. 以太网的 MAC 帧
72. 以太网的传输介质
73. 无线局域网
74. 广域网
75. 数据链路层设备
76. ★虚拟局域网（VLAN）
77. ★网络层
78. 子网划分和无分类编址 CIDR
79. 路由算法
80. IP 地址的分类、IP 数据报格式、NAT
81. ARP、DHCP 和 ICMP
82. 3 种常用路由选择协议：RIP、OSPF、BGP
83. IP 组播、移动 IP 的基本概念
84. 路由器的组成和功能
85. IPV6
86. 传输层
87. TCP 的流量控制和拥塞控制机制
88. ★TCP 报文格式
89. ★UDP 数据报格式
90. ★应用层
91. 域名解析过程
92. FTP 的工作原理
93. HTTP抓包与调试
94. Wireshark的使用与操作
95. 常见搜索引擎高级语法的使用
96. HTTP 请求与响应
97. HTTP 的请求方法
98. URL的特征分析
99. HTTP 消息头
100. HTTP状态码
101. 使用浏览器执行前端 JavaScript
102. 使用 Node.js 执行后端 JavaScript

第三编 数据科学基础知识

1. ★★Python编程基础
2. 基础语法与概念
3. 基本数据类型（整型、浮点型、字符串、布尔型）
4. 变量与赋值
5. 输入与输出（print函数和input函数）
6. 控制结构
7. 条件语句（if, elif, else）
8. 循环结构（while循环，for循环）
9. 循环控制语句（break, continue）
10. 简单列表推导式
11. 数据结构

列表（List）、元组（Tuple）、字典（Dictionary）、集合（Set）、数据结构的相关操作和方法

1. 函数与模块
2. 函数定义与调用
3. 参数传递与返回值
4. 局部变量与全局变量
5. 匿名函数（lambda表达式）
6. 模块的使用（import语句）
7. 常用标准库模块（如math, datetime）
8. 面向对象编程
9. 类与对象的概念
10. 类的定义与对象的创建
11. 实例属性与方法
12. 类属性与类方法
13. 继承与多态
14. 特殊方法（如\_\_init\_\_, \_\_str\_\_等）
15. 异常处理
16. 异常的概念
17. try-except语句
18. 处理多种异常
19. finally子句
20. 自定义异常
21. 文件与输入输出
22. 文件的打开与关闭
23. 读取与写入文件
24. 文件上下文管理器（with语句）
25. 文件的其他操作（如删除、重命名）
26. 高级主题
27. 上下文管理器与with语句
28. 正则表达式和GREP语法
29. 数据科学相关库（如NumPy, Pandas）
30. ★数据结构与算法基础
31. 二叉树的定义及其主要特征
32. 二叉排序树
33. 赫夫曼（Hoffman）树和赫夫曼编码
34. 树和森林的遍历
35. 图的基本概念
36. 图的遍历算法
37. 普里姆算法和克鲁斯卡尔算法
38. 迪杰斯特拉算法
39. 弗洛伊德算法
40. 关键路径核心算法
41. 拓扑排序核心算法
42. 排序算法
43. 顺序查找算法
44. 折半查找算法
45. 二排序树的基本算法
46. B-树的基本概念和操作
47. B+树的基本概念和操作
48. ★数据库管理与SQL
49. 数据库系统的基本概念
50. 数据库的分类和常见数据库软件
51. SQL 的特点
52. ★★使用 SQL 语言完成对数据库的查询、插入、删除、更新操作，可以用SQL语言正确完成复杂查询操作。
53. 查询多个列
54. 查询去重
55. SQL聚合函数
56. SQL 字符串函数
57. SQL数值函数
58. SQL⽇期和时间函数
59. SQL数据分组
60. ★★常见数据库的备份操作
61. ★★通过数据库日志恢复数据
62. ★常见数据库的客户端工具的使用
63. ★★能够在MySQL、SQLite上通过编程的方式开发应用程序，完成对数据库的各种操作；能够使用 ODBC、JDBC进行数据库应用程序的设计。
64. ★数据库连接地址、连接口令的分析方法
65. 数据处理
66. 数据清洗与预处理
67. ★使用Python对乱码，错位，重复值，未匹配等脏数据进行补全、删除等一致化处理。
68. 数据转换与特征工程常见方法
69. 缺失数据处理
70. 异常值检测与处理
71. ★★使用EXCEL或者WPS进行数据分析
72. 单元格格式调整方法
73. 数据查找和替换方法
74. 外部数据导入导出方法
75. 数据有效性检查方法
76. 日期和时间格式的处理方法
77. 数据筛选和排序
78. 数据分类汇总
79. 数据透视表、图的使用，包括分类汇总、取平均、最大最小值、自动排序、自动筛选、自动分组；可分析占比、同比、环比、定比
80. EXCEL两表数据比对方法
81. EXCEL常用统计公式使用
82. 日期函数：day，month，year，date，today，weekday，weeknum
83. 数学函数：product，rand，randbetween，round，sum，sumif，sumifs，sumproduct
84. 统计函数：large，small，max，min，median，mode，rank，count，countif，countifs，average，averageif
85. 查找和引用函数：choose，match，index，indirect，column，row，vlookup，hlookup，lookup，offset，getpivotdata
86. 文本函数：find，search，text，value，concatenate，left，right，mid，len
87. 逻辑函数：and，or，false，true，if，iferror
88. ★★文本编辑器数据处理
89. 列编辑模式的使用
90. 各类编码字符的搜索和替换
91. 正则搜索语法
92. 数据可视化
93. 数据可视化原理与技术
94. 使用Python/R进行数据可视化
95. 交互式数据可视化工具

第四编 机器学习与建模基础知识

1. 机器学习基础
2. 监督学习、无监督学习、半监督学习、强化学习
3. 机器学习算法概述
4. 模型评估与选择
5. 训练集与测试集划分
6. 交叉验证
7. 性能指标（精确度、召回率、F1分数、ROC曲线等）
8. ★机器学习算法
9. 线性回归与逻辑回归
10. 决策树与随机森林
11. 支持向量机（SVM）
12. 聚类分析与降维技术
13. ★深度学习基础
14. 神经网络基础
15. 前向传播与反向传播
16. 卷积神经网络（CNN）
17. 循环神经网络（RNN）
18. 模型融合与集成学习：bagging、boosting和stacking技术。
19. 时间序列分析：ARIMA、季节性分解与状态空间模型。
20. 自然语言处理（NLP）：词袋模型、TF-IDF、word2vec。
21. 推荐系统：协同过滤、基于内容的推荐、矩阵分解技术。
22. 模型部署与维护
23. 模型优化与调参
24. 模型持久化与版本控制
25. 在线与离线模型部署
26. 模型监控与维护

第五编 大数据技术基础知识

1. 大数据生态系统
2. 大数据定义与特征
3. Hadoop生态系统概述
4. 数据湖与数据仓库
5. ★Kubernetes（简称K8s）容器集群管理系统的安装部署
6. ★数据存储与计算
7. 分布式文件系统（HDFS）
8. NoSQL数据库（MongoDB, Cassandra）
9. 分布式计算框架（Spark, Flink）
10. 数据流处理
11. 实时数据处理概念
12. Kafka流处理平台
13. 流处理架构与应用

第六编 项目管理与业务理解基础知识

1. 数据项目管理
2. 敏捷项目管理方法
3. 数据项目生命周期
4. 团队协作与沟通技巧
5. 业务理解与分析
6. 业务需求分析
7. 数据驱动的决策制定
8. 伦理、法律与合规性
9. 数据伦理原则
10. 合规性检查与风险评估

第七编 网络安全基础知识

1. ★网络安全基础
2. 流量分析识别网络安全威胁
3. 端口扫描与监听
4. 口令破解
5. 拒绝服务攻击
6. 漏洞攻击（缓冲区溢出、系统漏洞、应用软件漏洞）
7. 僵尸网络
8. 网络钓鱼
9. 网络欺骗
10. 网站安全威胁的识别（SQL注入漏洞、跨域攻击、命令执行漏洞、文件漏洞等）
11. 常见的漏洞原理和利用方式，会使用工具或技术方法检测。
12. 恶意代码、文件传播方式判断和路径追溯
13. 通过Apache、Nginx、IIS等日志分析网络入侵行为
14. 利用系统日志、应用程序、安全日志等排查和溯源攻击行为。

四、笔试题型样例

（一）单选题

计算机中用来表示存储容量的基本单位是：

A. 字节

B. 字

C. 扇区

D. 磁道

（二）多选题

关于卷积神经网络（CNN），以下哪些说法是正确的？

A. CNN主要用于处理图像数据

B. CNN中的卷积层用于提取特征

C. CNN中的池化层用于减少参数数量

D. CNN不能用于处理序列数据，如文本或时间序列

（三）不定项选择题

根据《中华人民共和国个人信息保护法》，以下哪些行为是合法的？

A. 某电商网站在未经用户同意的情况下，收集用户的个人信息用于广告推送

B. 某银行为了防范电信诈骗，在未经用户同意的情况下，收集用户的通话记录和短信内容

C. 某医疗机构为了研究疾病，在经过用户同意的情况下，收集用户的健康信息

D. 某政府机构为了管理公共事务，在经过用户同意的情况下，收集用户的个人信息

E. 某社交媒体平台在用户注销账号后，仍然保留用户的个人信息

（四）代码阅读题

阅读以下Python代码，并描述这段代码的功能是什么？count\_chars("hello") 的返回值是什么？

def count\_chars(s):

count = {}

for char in s:

if char in count:

count[char] += 1

else:

count[char] = 1

return count

print(count\_chars("hello"))

（五）简答题

1. 请简述数据项目管理的主要任务是什么？

2. 请解释监督学习的概念，并列出至少3个常见的监督学习算法。

3. 如何实施有效的信息网络安全防护措施？

五、机试题型样例

（一）网络流量分析

对所给的网络数据包文件misc.pcapng进行分析，按要求回答问题。

1.查找数据包中异常的上传文件操作，列出上传的文件名。

2.文件ctf1.rar文件是通过什么方式发送的，请描述该文件的传播过程。

3.对ctf1.rar进行解密，展示解密后的文件ctf2.txt中的内容。

4.在查找数据包中查找webshell，并溯源分析该webshell的操作记录，描述该黑客的入侵过程。

（二）数据库分析

某市公安局派出所准备对辖区城中村开展治安清查行动，被清查人员中有“潘多拉”传销组织成员，被清查人员证件信息已导入数据库的sfzbiao3表，请结合数据库中的user\_biao1（传销组织成员表），money\_biao2（交易流水表）开展核查工作。所有数据存于SQLite数据库，文件名为“SQLite数据库.db”。

★提示：本题需使用数据库中“sfzbiao3”“money\_biao2”“user\_biao1”三个表完成操作。user\_biao1与money\_biao2通过字段Userid关联，层级字段为vip。★

请使用SQL数据库分析挖掘数据规律，并回答下列问题。

1.计算传销组织中各层级的人数，请在下面写出SQL语句，并将查询结果截图展示。

2.查询传销组织中广西籍人员，请在下面写出SQL语句，并将查询结果截图展示。

★提示：身份证号前两位45表示广西。

3.计算每个被清查人员的交易金额，请在下面写出SQL语句，并将查询结果截图展示。