

年 度	2022
编 号	QGB202203

2022年度“强国杯”技术技能大赛

——大数据技术应用赛项

技术方案

2022 年 06 月

目 录

一、 大赛名称.....	1
二、 大赛意义.....	1
三、 大赛内容、形式和成绩计算.....	1
(一) 竞赛内容.....	1
(二) 竞赛形式.....	2
(三) 参赛对象.....	2
(四) 报名条件.....	2
(五) 成绩计算.....	3
四、 奖励办法.....	4
五、 命题范围、赛题类型和其他.....	5
(一) 命题原则.....	5
(二) 考核范围、题型、考核时间.....	5
(三) 考核方式.....	8
(四) 命题方式.....	8
六、 大赛场地与设施.....	9
(一) 大赛场地.....	9
(二) 大赛设施.....	10
七、 大赛议程与时间安排.....	13
(一) 竞赛议程、晋级名额、议程总体安排.....	13
(二) 全国总决赛关键环节.....	14
(三) 全国总决赛竞赛流程.....	15

(四) 全国总决赛详细时间安排	15
八、 大赛赛题	16
九、 大赛评分标准制定原则、评分方法、评分细则及技术规范	16
(一) 评分标准制定原则	16
(二) 评分方法	16
(三) 评分细则	17
(四) 评分方式	20
(五) 技术规范	20
十、 大赛平台说明	21
(一) 硬件平台	21
(二) 软件设备	21
十一、 大赛安全保障	22
(一) 大赛保障与应急预案	22
(二) 操作与信息安全保障	23
十二、 大赛组织与管理	23
(一) 大赛设备与设施管理	23
(二) 大赛监督与仲裁管理	24
十三、 裁判人员要求	25
十四、 疫情防控	26

一、大赛名称

2022年度“强国杯”技术技能大赛——大数据技术应用赛项

官网：www.bigdata.qiangguobei.org.cn

二、大赛意义

2022年度“强国杯”全国大数据技术应用竞赛以习近平新时代中国特色社会主义思想为指导，深入贯彻落实习近平总书记对技能人才工作的系列重要指示，大力培育支撑中国制造、中国创造的高技术技能人才队伍，加快培养和选拔大数据领域高技术技能人才，弘扬精益求精的工匠精神，着力提高职业技能大赛科学化、规范化、专业化水平，坚持开放、公平、绿色、廉洁的办赛理念，创新大赛形式、提高大赛质量、推广大赛成果，实现以赛促学、以赛促教、以赛促建，为全面提高劳动者素质、推动经济高质量发展提供坚实基础，营造劳动光荣、技能宝贵、创造伟大的社会风尚，为全面建设社会主义现代化国家提供有力的大数据人才保障。

通过赛项的设立，加强专业人才培养，创新人才培养模式，建立健全多层次、多类型的大数据人才培养体系，采取产教融合等培养方式开展大数据综合型人才培养，鼓励企业合作和高等院校、职业院校，加强职业技能人才实践培养，积极培育大数据技术和应用创新型人才。通过赛项的设立，带动大数据相关人才的培养与专业发展，依托社会化教育资源，开展大数据知识普及和教育培训，提高社会整体认知和应用水平。

三、大赛内容、形式和成绩计算

（一）竞赛内容

2022年度“强国杯”全国大数据技术应用竞赛依据《大数据工程技术人员国家职业技术技能标准》，围绕大数据技术应用技能要求，参考国家职业技能大赛大数据工程技术人员标准进行赛题设计，内容以大数据技术应

用为核心，涵盖大数据环境部署与应用、大数据处理与应用、大数据分析与应用，融合项目案例，以理论为辅、实操为主综合考核选手大数据技术应用能力。

竞赛分为资格赛、半决赛、全国总决赛，其中资格赛、半决赛均包括理论和实操题目，资格赛以大数据主要分析应用工具为主，包括SQL语句使用、Python数据分析、大数据集群搭建、集群的应用、数据分析与挖掘工具应用。半决赛以大数据集群运维与应用、数据采集与处理、数据可视化、Spark MLlib应用等内容为主。全国总决赛以案例实战为主，综合考核大数据集群优化、数据任务调度应用、自然语言处理和机器学习。

（二）竞赛形式

竞赛以个人赛形式开展，采用理论+实操形式，考核选手的大数据技术应用综合素养。

（三）参赛对象

职工选手：从事大数据应用等相关专业或职业的企业在职职工、各类高等院校和职业院校（含技工院校、高职高专、中职等，下同）教职工人员。

本科学生选手：各类本科类型院校全日制在籍学生，每位学生可报一名指导教师，指导教师不参加竞赛。

职教学生选手：各类职业教育类型院校（含技师院校）全日制在籍学生，每位学生可报一名指导教师，指导教师不参加竞赛。

（四）报名条件

各参赛组别具体参赛对象要求如下：

职工组：从事计算机运维、软件开发、大数据技术、人工智能技术、云计算技术、物联网技术、网络安全技术相关工作的企业在职人员、学校

全职教职工人员。非在职人员，不得以职工选手身份报名。职工选手请选择“职工组”报名，一经报名不得修改，报名时须按报名要求提交个人信息及社保证明，无法提供社保证明的须提交单位推荐信报名参加，每家单位报名人数不限。

本科学生组:计算机、软件工程、大数据、人工智能、云计算、物联网、网络安全等相关专业的本科类型全日制在校学生，选手可配备一名指导教师，指导教师不参与竞赛。本科学生选手请选择“本科学生组”报名。报名时须按要求提交报名信息，提交身份证复印件、学生证复印或二级学院（系）推荐信，每所学校报名人数不限。

职教学生组：计算机、软件工程、大数据、人工智能、云计算、物联网、网络安全等相关专业的职业教育类型全日制在校学生，选手配备一名指导教师，指导教师不参与竞赛。职教学生选手请选择“职教学生组”报名。报名时须按要求提交报名信息，提交身份证复印件、学生证复印或二级学院（系）推荐信，每所学校报名人数不限。

（五）成绩计算

（1）资格赛阶段

资格赛阶段，理论题目占比为40%，赛题均为客观题，采用机考方式实现；实操题目占比为60%，赛题均为实操题，采用机考方式实现。理论题目、实操题目将按比例折算计入资格赛总成绩。

（2）半决赛阶段

半决赛阶段，理论题目占比为20%，赛题均为客观题，采用机考方式实现；实操题目占比为80%，赛题均为实操题，采用机考方式实现。理论题目、实操题目将按比例折算计入半决赛总成绩。

（3）全国总决赛阶段

全国总决赛阶段，赛题均为实操题，采用机考方式实现，由系统根据选手操作情况智能检测评分。

四、奖励办法

(1) 半决赛职工组、本科学生组、职教学生组各设半决赛一、二、三等奖奖项，以实际参赛人数为基数，一、二、三等奖获奖比例分别为10%、20%、30%，获奖选手由竞赛技术委员会（北京红亚华宇科技有限公司代章）颁发半决赛荣誉证书（电子版）。

(2) 全国总决赛职工组、本科学生组、职教学生组各组成成绩第一名者获得金牌奖、成绩第二名者获得银牌奖、成绩第三名者获得铜牌奖。并可获得对应奖项奖杯。各组金奖选手另外可获奖金3万元（税前），银奖选手另外可获得奖金2万元（税前），铜奖选手另外可获奖金1万元（税前）。

(3) 全国总决赛职工组、本科学生组、职教学生组各设一、二、三等奖奖项，以总决赛实际参赛人数为基数，一、二、三等奖获奖比例分别为10%、20%、30%，获奖选手由主办单位颁发奖项荣誉证书。一等奖另可获得由北京红亚华宇科技有限公司组织的“大数据技术应用工程师（高级）”免费培训券一张（价值4400元）。二等、三等奖可获得由北京红亚华宇科技有限公司组织“大数据技术应用工程师（中级、高级）”培训券一张（价值2000元）。工业和信息化部教育考试中心向参训并考试合格学员颁发职业能力证书（中级或高级），学员信息将纳入“工业和信息化技术技能人才库”，可在官网（www.miiteec.org.cn）查询。

(4) 全国总决赛学生组一等奖获奖队伍的教练（指导教师），颁发“优秀教练”证书。

(5) 对全国总决赛贡献突出的承办、协办和技术支持单位，颁发

“突出贡献单位”奖牌和证书。

(6) 对全国总决赛竞赛组织实施中表现突出的个人，颁发“优秀工作者”证书。

(7) 对在全国总决赛执裁工作中表现突出的裁判，颁发“优秀裁判员”证书。

五、命题范围、赛题类型和其他

(一) 命题原则

此次竞赛设计完全与工作实需求为背景，按照大数据工程技术员国家职业技能标准，聚焦大数据技术应用技术前沿趋势，围绕大数据环境部署与应用、数据采集与治理、大数据分析与挖掘、数据分析算法与数据可视化技术应用等方面进行重点考察，突出企业所需专业技能及新技术应用，遵循竞赛一贯坚持的公开、公平、公正原则，突出职业能力考核及工匠精神要求。

(二) 考核范围、题型、考核时间

阶段	时长	模块	模块名称及考核内容
资格 赛	120 分 钟	A	理论知识模块 <ul style="list-style-type: none">• 分析工具知识• 数据库知识• 数据分析知识• 数据管理知识• 数据安全知识• 数据保密知识• 相关法律知识
		B	EXCEL/WPS 工具数据分析 <ul style="list-style-type: none">• 工具设置及快捷键使用• 数据分列• 数据筛选• 统计分析函数/文本处理函数• 数值运算函数/逻辑判断函数

半决赛	180分钟		<ul style="list-style-type: none"> 日期计算函数/匹配查找函数 使用函数实现报表开发 数据透视基础表制作
		C	SQL 数据分析
			<ul style="list-style-type: none"> DDL/DML/DQL/DCL 使用 SQL 查询关键词 SQL 核心框架 SQL 书写顺序 简单查询 条件查询 多表查询
			Python 数据分析
		D	<ul style="list-style-type: none"> Python 基础语法使用 Numpy 数据计算 Pandas 数据分析 Matplotlib 数据可视化
			伪分布式集群搭建
		E	<ul style="list-style-type: none"> Linux 操作系统 主机名/防火墙设置 系统服务开启/关闭 Unix 系统文本编辑器使用 SSH 远程登录 JDK 安装部署 Hadoop 伪分布式部署
			理论知识模块
			F
		大数据平台运维	
G	<ul style="list-style-type: none"> 集群配置与管理 大数据各组件常见参数配置 		

			<ul style="list-style-type: none"> 查看组件相关日志 精准故障判断 标准化运维维修 集群本地测试
		H	数据迁移 <ul style="list-style-type: none"> 数据传输工具 Sqoop 环境部署 传统数据库数据导入大数据集群 大数据集群数据出到传统数据库
		I	数据采集 <ul style="list-style-type: none"> 使用爬虫代码采集数据 使用 Flume 进行日志采集
		J	离线数据处理 <ul style="list-style-type: none"> 编写 MapReduce 进行数据分析 使用 Hive SQL 进行数据分析
		K	实时数据处理 <ul style="list-style-type: none"> Kafka 集群部署 使用 Kafka 消费数据 使用 Spark 进行数据处理 使用 HBase 进行实时读写
		L	数据可视化 <ul style="list-style-type: none"> 基于前端进行数据可视化展示 使用工具进行数据可视化展示
		M	Spark MLlib <ul style="list-style-type: none"> 基本统计分析 特征抽取 特征变换 分类与回归
全国总决赛	240 分钟	N	集群优化 <ul style="list-style-type: none"> 操作系统调优 HDFS 参数调优 MR 参数调优 YARN 参数调优 应用系统调优 集群数据均衡 集群读写性能测试

		O	开源任务调度系统应用
			<ul style="list-style-type: none"> • workflow 调度实现方式 • workflow 定义 • workflow 传参 • 定时执行 • 资源管理 • workflow 执行 • workflow 管理
		P	运用 Python 编写算法对文本类数据案例分析
			<ul style="list-style-type: none"> • 文本数据获取 • 分词/词向量 • 构造词频矩阵 • 逻辑回归/贝叶斯分类 • 分类结果可视化
		Q	运用 Python 编写算法对图像类数据案例分析
			<ul style="list-style-type: none"> • 图像采集 • 图像增强 • 灰度/二值化处理 • 边缘/轮廓检测 • 傅立叶变换/霍夫变换 • 目标识别

(三) 考核方式

竞赛各赛段要求以公开、公平、公正原则开展，采用系统智能评测，综合考虑时间因素，竞赛环节中实时公开成绩，公开选手动态，实时公开选手排行榜。考核方式采用先进的智能化测评技术，将所有竞赛结果进行智能评测，全面体现公开、公平、公正原则，通过云答题平台的创新的自动化监测技术，自动识别参赛选手的平台搭建是否成功、自动判断数据分析结果、自动评估算法模型识别精度等，全程采用自动化智能判断。

(四) 命题方式

本竞赛命题是从实际企业项目出发，将基础技能和工作任务融合的方式命题。赛项题目的设计，聚焦大数据技术应用技术前沿趋势，围绕大数

据环境部署与应用、数据采集与治理、大数据分析与挖掘、数据分析算法与数据可视化技术应用等方面进行重点测评，考察选手的综合应用能力。

六、大赛场地与设施

(一) 大赛场地

竞赛现场设置场内竞赛区、现场裁判工作区、技术支持区，场外互动区等。

监考人员协助裁判长和现场裁判做好负责工位范围内的秩序维持，监考人员不得在考场内随意走动。

技术支持保障人员在技术支持服务区候场，有需要时在现场裁判的带领下到相关的工位进行赛场技术支持保障，在条件具备时，技术支持区可不设置在考场内。

独立工位的竞赛场地需求至少为6平方米，竞赛现场符合消防安全规定，现场消防器材和消防栓合格有效，应急照明设施状态合格，赛场明显位置张贴紧急疏散图，赛场地面张贴荧光疏散指示箭头，赛场出入口专人负责，随时保证安全通道的畅通无阻。

各工位分区供电，强电弱电分开布线，工位及竞赛桌面照度大于500lux。现场临时用电需满足《施工现场临时用电安全技术规范》JGJ46-2005的要求。

竞赛现场需通风良好、照明需符合教室采光规范。

竞赛场地初步按照可容纳500人的规模设计，并视最终报名情况，及时调整场地布置。

(二) 大赛设施

1.比赛器材、技术平台

(1) 硬件设备

设备类别	数量	基本配置
大数据竞赛云平台	1套	内嵌大数据竞赛管理系统。提供大数据竞赛、虚拟机操作、大数据竞赛管理等服务。
客户端	500台	CPU性能相当于2.0GHZ处理器，8G以上内存，500GB以上硬盘，显示器要求1920*1080以上。

注：视场地情况，由竞赛承办单位统一提供，如场景无法统一提供情况下，须由选手自备电脑，具体以决赛通知为准。

(2) 软件系统（大数据竞赛管理系统一套）

大数据竞赛竞赛系统包含本次竞赛所需要的竞赛环境、安装软件、其他工具等。支持各种类型数据的存储和组织，支持多种类型数据的查询、统计分析处理要求；支持并行SQL计算框架、流计算框架、内存计算框架；支持对集群计算资源和处理作业进行统一分配和调度；提供对实时、近实时和离线等多种场景的支持；提供大数据的采集、提取、清洗、转换、分析操作，产生分析结果。

系统应结合大数据集群搭建与运维所需软件资源包、部署过程等，为每个阶段的设定自动检测方法，包括文件位置检测、文件内容检测、系统服务检测、系统配置检测等，通过检测每道题目的答案是否与标准答案相匹配、以及结合过程性的AI智能分析技术，自动检测选手提交的测试结果，并通过实时计算实时展示队伍成绩。

支持进行多种类型的大数据相关内容的训练及比赛，支持在线进行操作系统级的真机模拟训练测试，充分体验和掌握大数据相关技能。支持组织和安排大数据竞赛，提供实时了解用户竞赛状态的监控能力。比赛任务

完成后，支持通过查看用户的测评报告了解用户对大数据相关知识点和操作技能的掌握情况。

设备类别	软件类别	软件名称
服务端	大数据竞赛系统运行操作系统	Centos7.0
	大数据平台软件	大数据竞赛系统
	数据库	Mysql5.7
	网站开发语言	Php 7.0
	消息队列程序	Redis3
客户端	桌面操作系统	Windows10
	浏览器	Chrome64
	文档编辑器	Office 2007 及以上版本

大学生大数据竞赛

总倒计时: 02:32:32

年轻人不讲武德 - Alina 暂时退出 结束答题

阶段安排

- 1 上机实验 开始时间: 2021-12-25 19:23 倒计时: 12:32:12
- 2 理论评测 开始时间: 2021-12-25 19:23 倒计时: 12:32:12
- 3 上机实验 开始时间: 2021-12-25 19:23 倒计时: 12:32:12

操作环境 | 通知栏 | 排行榜 | 竞赛信息

公告 2021-12-25 19:23:42

该调色板包含了主色和强调色，可以用于插图或开发你的品牌色。这些经过设计的颜色可以灵活的搭配。调色板从主色开始，通过填充光谱为 Android、Web、iOS 创建了完整可用的调色板。Google 建议使用 500 色号的颜色作为主色，其他颜色作为强调色。

公告 2021-12-25 19:23:42

该调色板包含了主色和强调色，可以用于插图或开发你的品牌色。这些经过设计的颜色可以灵活的搭配。调色板从主色开始

题目预览页

演示赛 总倒计时: 02:38:15 张程 暂时退出

阶段安排

- 1 实操模块 开始时间: 2022-01-05 16:13:08 倒计时: 02:51:21
- 2 大数据理论模块 开始时间: 2022-01-05 16:13:08 倒计时: 02:51:21

操作环境 | 通知栏 | 排行榜 | 竞赛信息

slave1 启动

direct_ssh
操作环境类型: VM
用户名: -
密码: -
端口: -
IP: -
Public_IP: -

master 启动

direct_ssh
操作环境类型: VM
用户名: -
密码: -
端口: -

返回答题情况

模块说明

1

题目一、基础配置 (0 / 100分)

1.比赛框架
本次比赛为分布式集群搭建, 共三台节点, 其中master作为主节点, slave1、slave2为从节点;

2.比赛内容

- 基础配置: 修改主机名、主机映射、时区修改、时间同步、定时任务、免密访问;
- JDK安装: 环境变量;
- Zookeeper部署: 环境变量、配置文件zoo.cfg、myid;
- Hadoop部署: 环境变量、配置文件修改、设置节点文件、格式化、开启集群;
- Hive部署: Mysql数据库配置、服务器端配置、客户端配置。

3.版本说明

内置安装/依赖包 (/usr/package277)	已安装服务	系统版本
hadoop-2.7.7.tar.gz	ntp	CentOS Linux release 7.3.1611 (Core)
zookeeper-3.4.14.tar.gz	mysql-community-server	
apache-hive-2.3.4-bin.tar.gz		
jdk-8u211-linux-x64.tar.gz		
mysql-connector-java-5.1.47-bin.jar		

core-site.xml参数配置详情

官方文档: <http://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-common/core-default.xml>

core-default.xml与core-site.xml的功能是一样的, 如果在core-site.xml里没有配置的属性, 则会自动会获取core-default.xml里的相同属性的值

属性	值	说明
fs.default.name	hdfs://????	定义master的URI和端口

选手答题页



选手成绩展示页

(3) 竞赛平台支持

平台支撑单位承诺为决赛提供免费平台支持，确保竞赛稳定、正常进行。

2.其他要求

防护设备：灭火器、供电系统。

禁止携带物品：移动存储、手机、上网卡等与竞赛无关附件。

七、大赛议程与时间安排

(一) 竞赛议程、晋级名额、议程总体安排

竞赛按照“公开、公平、公正”的原则开展竞赛活动，职工选手、本科学生选手、职教学生选手各组整体议程分为资格赛、半决赛和全国总决赛三个阶段，如对应组别内有单独竞赛要求，另行通知。各环节实施方案如下：

1.资格赛阶段

各组选手可积极报名参加资格赛，资格赛人数不限，采用线上方式进行，题目为理论与实践结合方式考核，理论占比40%，实操占比60%，竞赛时长120分钟，各组参赛选手获得资格赛题总分数的60%者，即可进入各组半决赛。

2022年8月31日前，在全国组委会的统一指导下，按照竞赛规程，完成各组选手资格赛。

2.半决赛阶段

各组选手半决赛采用线上方式进行，题目为理论与实践结合方式考核，理论占比20%，实操占比80%，竞赛时长180分钟，选手通过半决赛后，可进入各组全国总决赛。

职工选手总成绩排名前200名选手、本科学生选手总成绩排名前150名选手、职教学生选手总成绩排名前150名选手，进入各组全国总决赛。

2022年10月31日前，在全国组委会的统一指导下，按照竞赛规程，选拔优秀选手参加全国总决赛。

3.全国总决赛阶段

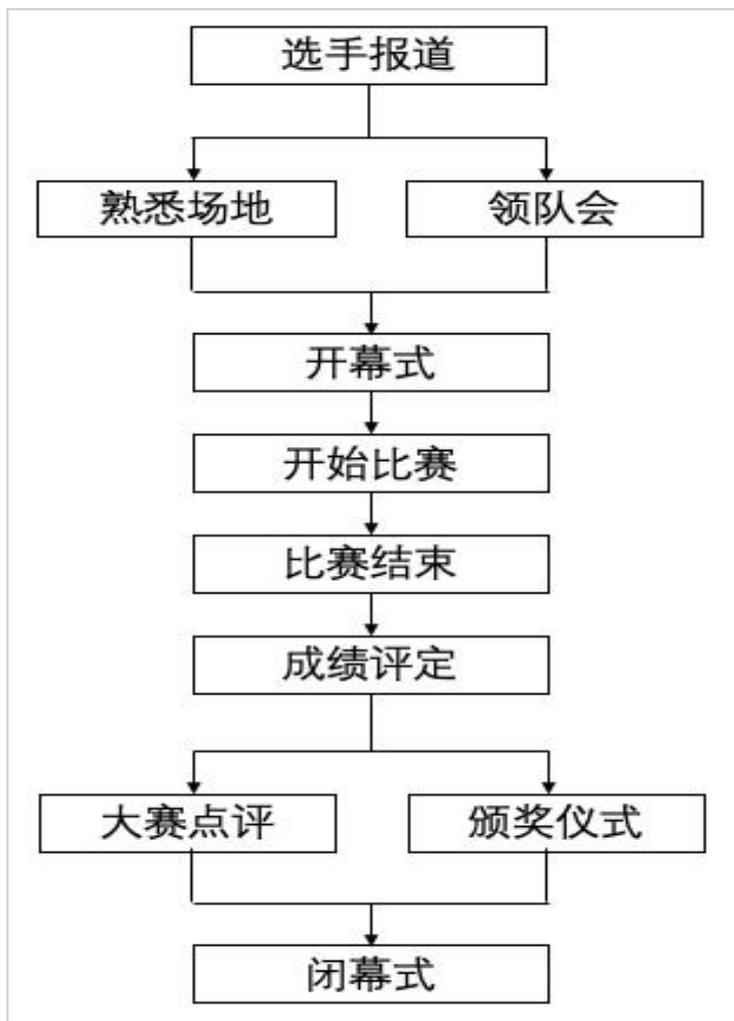
各组选手全国总决赛题目全部为实操题型，竞赛时长240分钟，根据按答题正确情况，结合时间因素进行排名，各组选手独立排名。

2022年12月31日前，组织完成各组全国总决赛。全国总决赛时间、地点与赛程安排另行通知。

(二) 全国总决赛关键环节

参赛选手报到——参赛选手赛前熟悉场地、领队会——开幕式——正式比赛——比赛结束——成绩评定——竞赛技术点评、颁奖 仪式、闭幕式（如疫情影响无法线下开展的，采用线上决赛，决赛流程分别通知）。

(三) 全国总决赛竞赛流程



(四) 全国总决赛详细时间安排

对接竞赛指南日程安排，竞赛时长4小时。具体日程安排详见下表。

日期	时间	事项	参加人员	地点
报道日	08:00-13:00	职工组、本科学生组、 职教学生组报道		住宿酒店
	14:45-15:30	赛前说明会	参赛代表	会议室
	15:30-16:20	熟悉赛场	参赛代表	竞赛场地
竞赛日	07:30	参赛选手住宿宾馆门口 集合，集体乘车往赛场	参赛选手	住宿酒店
	08:30-9:00	领导致辞	领导、嘉宾、裁	竞赛场地

			判、参赛选手	
	09:00-13:00	竞赛	参赛选手	竞赛场地
	13:00-14:30	午餐	参赛选手、裁判、工作人员	竞赛场地
	14:30-15:30	闭赛式	领导、嘉宾、裁判长、裁判、参赛选手	报告厅

八、大赛赛题

大赛组委会拟在 2022年6月份组织技术说明会。2022年7月在大赛官方网站上发布全国总决赛比赛样题及竞赛所使用关键部件使用手册(网址为：www.bigdata.qiangguobei.org.cn)。

九、大赛评分标准制定原则、评分方法、评分细则及技术规范

(一) 评分标准制定原则

依据参赛选手完成的情况实施综合评定。评定依据2022年度“强国杯”技术技能大赛—大数据技术应用赛项竞赛实施方案中明确的技术规范，按照技能竞赛技术裁判组制定的考核标准进行评分，全面评价参赛选手职业能力的要求，本着“科学严谨、公开公平公正、可操作性强、突出工匠精神”的原则制定评分标准。

(二) 评分方法

1.基本评定方法

裁判组在坚持“公开、公平、公正”的原则下，各负其责，按照制订的评分细则进行评分，赛项评分采用结果评分方法。结果评分：依据赛项评价标准，对参赛选手提交的竞赛结果进行评分，赛项最终按总评分得分高低排名。

成绩汇总：比赛成绩由竞赛系统智能评测，评测完成后，经过评分裁判审核无异议后，由仲裁组长复核后确认并发布。

2.相同成绩处理

总成绩相同时，以实操总成绩得分高的名次在前；总成绩和实操比赛总成绩相同时，系统工作运行效率得分高的名次在前。

(三) 评分细则

阶段	时长	模块	模块名称及考核内容	分值占比	评分方式
资格赛	120分钟	A	理论知识模块	40%	系统自动评分
			<ul style="list-style-type: none"> 分析工具知识 数据库知识 数据分析知识 数据管理知识 数据安全知识 数据保密知识 相关法律知识 		
		B	EXCEL/WPS 工具数据分析	10%	系统自动评分
			<ul style="list-style-type: none"> 工具设置及快捷键使用 数据分列 数据筛选 统计分析函数/文本处理函数 数值运算函数/逻辑判断函数 日期计算函数/匹配查找函数 使用函数实现报表开发 数据透视基础表制作 		
C	SQL 数据分析	15%	系统自动评分		
	<ul style="list-style-type: none"> DDL/DML/DQL/DCL 使用 SQL 查询关键词 SQL 核心框架 SQL 书写顺序 简单查询 条件查询 多表查询 				
D	Python 数据分析	15%	系统自动		

			<ul style="list-style-type: none"> Python 基础语法使用 Numpy 数据计算 Pandas 数据分析 Matplotlib 数据可视化 		评分
		E	伪分布式集群搭建 <ul style="list-style-type: none"> Linux 操作系统 主机名/防火墙设置 系统服务开启/关闭 Unix 系统文本编辑器使用 SSH 远程登录 JDK 安装部署 Hadoop 伪分布式部署 	20%	系统自动评分
半决赛	180 分钟	F	理论知识模块 <ul style="list-style-type: none"> 大数据基本概念 Hadoop 基础知识 Hive 基础知识 Spark 生态体系 数据迁移知识 数据采集知识 离线处理与实时处理 大数据平台运维基础知识 	20%	系统自动评分
		G	大数据平台运维 <ul style="list-style-type: none"> 集群配置与管理 大数据各组件常见参数配置 查看组件相关日志 精准故障判断 标准化运维维修 集群本地测试 	10%	系统自动评分
		H	数据迁移 <ul style="list-style-type: none"> 数据传输工具 Sqoop 环境部署 传统数据库数据导入大数据集群 大数据集群数据出到传统数据库 	10%	系统自动评分
		I	数据采集 <ul style="list-style-type: none"> 使用爬虫代码采集数据 使用 Flume 进行日志采集 	10%	系统自动评分
		J	离线数据处理	15%	系统自动

			<ul style="list-style-type: none"> • 编写 Mapreduce 进行数据分析 • 使用 Hive SQL 进行数据分析 		评分
		K	实时数据处理 <ul style="list-style-type: none"> • Kafka 集群部署 • 使用 Kafka 消费数据 • 使用 Spark 进行数据处理 • 使用 HBase 进行实时读写 	15%	系统自动评分
			数据可视化 <ul style="list-style-type: none"> • 基于前端进行数据可视化展示 • 使用工具进行数据可视化展示 		
		M	Spark MLlib <ul style="list-style-type: none"> • 基本统计分析 • 特征抽取 • 特征变换 • 分类与回归 	15%	系统自动评分
全国总决赛	240 分钟	N	集群优化 <ul style="list-style-type: none"> • 操作系统调优 • HDFS 参数调优 • MR 参数调优 • YARN 参数调优 • 应用系统调优 • 集群数据均衡 • 集群读写性能测试 	20%	系统自动评分
			开源任务调度系统应用 <ul style="list-style-type: none"> • 工作流调度实现方式 • 工作流定义 • 工作流传参 • 定时执行 • 资源管理 • 工作流执行 • 工作流管理 		
		P	运用 Python 编写算法对文本类数据案例分析 <ul style="list-style-type: none"> • 文本数据获取 • 分词/词向量 • 构造词频矩阵 	30%	系统自动评分

		<ul style="list-style-type: none"> • 逻辑回归/贝叶斯分类 • 分类结果可视化 		
	Q	运用 Python 编写算法对图像类数据案例分析	30%	系统自动评分
		<ul style="list-style-type: none"> • 图像采集 • 图像增强 • 灰度/二值化处理 • 边缘/轮廓检测 • 傅立叶变换/霍夫变换 • 目标识别 		

(四) 评分方式

竞赛平台内置自主检测功能，选手提交检测后，系统将考核要求对应的智能检测方法与实际操作结果自动化匹配，匹配成功取得分数，匹配失败不得分，完全采用客观化评分，评分项内无主观分值。

(五) 技术规范

本次评分规则参照世界技能大赛评分规则执行。本赛项评分标准为结果评分。采用客观数据结果测评的评判称为结果评分；结果评分方式：依照竞赛题库评价标准，对选手提交的答案进行自动测评，判定结果正确、准确度达标即可得到对应分值。

结果评分准则样例表：

类型	示例	最高分值	正确分值	不正确分值
从零分开始加	Print(“Welcome to the python match!”), 系统检测核心代码Print()及运行结果“Welcome to the python match!”是否与答案匹配，代码及结果完全全匹配得分，不匹配不得分。	题目规定分	满分	0

十、大赛平台说明

(一) 硬件平台

大数据竞赛云平台能够对平台的硬件资源、网络、镜像等进行管理调配，在使用过程中对每一名参赛选手所使用的资源均衡分配，同时保证了虚拟机高并发的稳定运行。

客户端为标准的PC，参赛选手使用浏览器软件访问竞赛平台，并使用连接工具连接虚拟机进行竞赛操作。

(二) 软件设备

大数据竞赛系统包含本次竞赛所需要的竞赛环境、安装软件、其他工具等。支持各种类型数据的存储和组织，支持多种类型数据的查询、统计分析处理要求；支持并行SQL计算框架、流计算框架、内存计算框架；支持对集群计算资源和处理作业进行统一分配和调度；提供对实时、近实时和离线等多种场景的支持；提供大数据的采集、提取、清洗、转换、分析操作，产生分析结果。

竞赛系统结合大数据技术应用题目所需的软件资源包、部署过程、数据分析过程等，为每个阶段的设定自动检测方法，包括文件位置检测、文件内容检测、系统服务检测、系统配置检测等，通过检测每道题目的答案是否与标准答案相匹配、以及结合过程性的AI智能分析技术，自动检测选手提交的竞赛结果，并通过实时计算实时展示队伍成绩。

支持进行多种类型的大数据相关内容的训练及比赛，支持在线进行操作系统级的真机模拟训练测试，充分体验和掌握大数据相关技能。支持组织和安排大数据竞赛，提供实时了解用户竞赛状态的监控能力。比赛任务完成后，支持通过查看用户的测评报告了解用户对大数据相关知识点和操作技能的掌握情况。

十一、大赛安全保障

(一) 大赛保障与应急预案

赛场组织与管理应制定安保须知、安全隐患规避方法及突发事件预案，设立紧急疏散路线及通道等，确保比赛期间所有进入竞赛地点的车辆、人员需凭证入内；严禁携带易燃易爆物、管制刀具等危险品及比赛严令禁止的其他物品进入场地；对于紧急发生的拥挤、踩踏、地震、火灾等进行紧急有效的处置。

1.环境保障

赛项执委会须在赛前组织专人对比赛现场、住宿场所和交通保障进行考察，并对安全工作提出明确要求。赛场的布置、赛场内的器材、设备，应符合国家有关安全规定。如有必要，也可进行赛场仿真模拟测试，以发现可能出现的问题，承办单位赛前须按照赛项执委会要求排除安全隐患。

防止无关人员发生意外事件，比赛现场内应参照相关职业岗位要求为选手提供必要的劳动保护，在具有危险性的操作环节，裁判员要严防选手出现错误操作。

2.应急预案实施的条件

对于比赛内容涉及高空作业、可能有坠物、大用电量、易发生火灾等情况的赛项，必须明确制度和预案，并配备急救人员与设施。

4.人员疏导方案

赛场环境中存在人员密集、车流人流交错的区域，除了设置齐全的指示标志外，须增加引导人员，并开辟备用通道。

5.外部因素

参赛选手进入赛位、赛事裁判工作人员进入工作场所，严禁携带U盘、手机等任何禁止携带的通讯、记录设备。如确有需要，由赛场统一配

置、统一管理。赛项可根据需要配置安检设备对进入赛场重要部位的人员进行安检。

(二) 操作与信息安全保障

安装UPS：采用UPS防止现场因突然断电导致的系统数据丢失，额定功率：3KVA，后备时间：2小时，电池类型：输出电压：230V±5%V；市电采用双路供电。

赛前要对选手进行计算机、网络设备、工具等操作的安全培训，进行安全操作的宣讲，确保每个队员能够安全操作设备后方可进行比赛。裁判员在比赛前，宣读安全注意事项，强调用火、用电安全规则。

竞赛期间，赛项承办单位须在赛场管理的关键岗位，增加力量，建立安全管理日志。

整个竞赛过程邀请当地公安系统、卫生系统和保险系统协助支持。

参赛队：由各省市负责参赛选手旅途及竞赛过程中的安全保障，应按要求为参赛选手购买保险。

十二、大赛组织与管理

(一) 大赛设备与设施管理

竞赛现场设置技术支持区，技术支持保障人员在技术支持服务区候场，有需要时在现场裁判的带领下到相关的工位进行赛场技术支持保障，同时大数据竞赛云平台放置在技术支持区，技术支持保障人员负责运维。

场内竞赛区为参赛队提供统一的竞赛设备；无需选手自带任何工具及附件。

独立工位的竞赛场地需求至少为6平方米，竞赛现场符合消防安全规定，现场消防器材和消防栓合格有效，应急照明设施状态合格，赛场明显位置张贴紧急疏散图，赛场地面张贴荧光疏散指示箭头，赛场出入口专人

负责，随时保证安全通道的畅通无阻。

各工位分区供电，强电弱电分开布线，工位及竞赛桌面照度大于500lux。现场临时用电需满足《施工现场临时用电安全技术规范》JGJ46-2005的要求。

竞赛现场需通风良好、照明需符合教室采光规范。

竞赛场地初步按照可容纳500人的规模设计，并视最终报名情况，及时调整场地布置。

（二）大赛监督与仲裁管理

（1）大赛采取二级仲裁机制。赛项设赛项仲裁工作组、大赛执委会设仲裁委员会。各参赛队对不符合大赛和赛项规程规定的仪器、设备、工装、材料、物件、计算机软硬件、竞赛使用工具、用品，竞赛执裁、赛场管理，以及工作人员的不规范行为等，可向赛项仲裁工作组提出申诉。申诉主体为参赛队领队。申诉启动时，领队向赛项仲裁工作组递交亲笔签字同意的书面申诉报告。申诉报告应对申诉事件的现象、发生时间、涉及人员、申诉依据等进行充分、实事求是的叙述。非书面申诉不予受理。

（2）申诉应在竞赛结束后1个小时内提出，超过时效将不予受理。申诉时，应按照规定程序由参赛队领队向相应赛项仲裁工作组递交书面申诉报告。报告应对申诉事件的现象、发生的时间、涉及到的人员、申诉依据与理由等进行充分、实事求是的叙述。事实依据不充分、仅凭主观臆断的申诉不予受理。申诉报告须由申诉的参赛选手、领队签名。

（3）赛项仲裁工作组收到申诉报告后，应根据申诉事由进行审查，2小时内通知申诉方，告知申诉处理结果。

（4）申诉人不得无故拒不接受处理结果，不允许采取过激行为刁难、攻击工作人员，否则视为放弃申诉。申诉人不满意赛项仲裁工作组的

处理结果的，可向大赛仲裁委员会提出复议申请。大赛仲裁委员会在接到复议申请后的1天内组织复议，并及时反馈复议结果，以大赛仲裁委员会的仲裁结果为最终结果。

十三、裁判人员要求

建议由本科高等院校、职业院校及行业、企业专家共同构成裁判组。对裁判组成员及数量的要求为：1名裁判长、3名裁判员、3名评分员，共7人。裁判长要求：身体健康，年龄一般在65周岁以下，具有良好的职业道德，坚持原则、作风正派、认真负责、廉洁公正，从事大数据、计算机、软件、网络、电子等专业工作或教学经验5年以上，有较深的理论造诣，熟悉本专业国内外的技术标准和业务流程，在全国专业领域内有一定的权威性和知名度，具有大数据相关的高级证书人员。

序号	专业技术方向	知识能力要求	执裁、教学、工作经历	专业技术职称（职业资格等级）	人数
1	大数据专业方向	熟练掌握大数据应用技术、大数据运维技术、大数据分析技术、大数据存储技术	五年及以上相关教学经验，具有省级或行业职业技能竞赛执裁经验	副教授（含副高级工程师）以上职称或获得过大数据技术相关的高级职业能力证书人员	4
2	计算机应用专业方向	熟悉操作系统、数据结构、数据库及常用的编程语言	五年及以上相关教学经验，具有省级或行业职业技能竞赛执裁经验	副教授（含副高级工程师）以上职称，获得过计算机应用相关的高级职业能力证书人员	1
3	网络工程专业方向	熟悉网络协议、常用网络设备 etc 知识	五年及以上相关教学经验，具有省级或行业职业技能竞赛执裁经验	副教授（含副高级工程师）以上职称获得过计算机网络相关的高级职业能力证书人员	1

4	人工智能专业方向	熟悉人工智能算法、深度学习、机器学习专业技术应用等知识	五年及以上相关教学经验，具有省级或行业职业技能竞赛执裁经验	副教授（含副高级工程师）以上职称或获得过大数据、人工智能相关的高级职业能力证书人员	1
裁判总人数	7				

十四、疫情防控

（1）请各相关单位高度重视疫情防控要求，按照属地要求，提前做好相关准备工作，确保大赛安全顺利进行。

（2）请各参赛队及各有关单位自大赛前第 14 天起，对所有参加大赛人员进行体温检测和健康状况监测。按照“异常人员应检尽检、其他人员愿检尽检”的原则，对身体状况出现异常和监测发现身体状况异常的人员进行核酸检测。

（3）请各参赛队及所有参加大赛人员出发前自行查验“一卡一码一证明”，即行程卡、健康码和核酸检测证明。低风险地区所有参加大赛人员需持健康通行码“绿码”，在测温正常且做好个人防护前提下可有序流动，进入密闭会场时需佩戴普通医用口罩。中、高风险地区所有参加大赛人员需持有抵达前7日内核酸检测阴性证明和健康通行码“绿码”，在测温正常且做好个人防护前提下可有序流动，进入密闭会场时需佩戴普通医用口罩。

（4）所有参加大赛人员体温低于37.3℃方可入场。身体状况异常的，大赛承办单位将协调卫生健康部门组织疾控机构和医疗机构专家对其进行核酸检测，并提出专业评估建议。

（5）疫情防控其他未尽事宜按属地疫情防控政策执行。

