

辽宁省职工技能大赛工作领导小组办公室

辽职赛办发〔2022〕44号

关于举办 2022 辽宁省职工技能大赛 暨第二届全省大数据应用 技能大赛的通知

各拟参赛单位：

按照辽宁省总工会、辽宁省人力资源和社会保障厅联合下发的《关于组织开展 2022 辽宁省职工技能大赛的通知》（辽工发〔2022〕19号）安排，切实组织好 2022 辽宁省职工技能大赛暨第二届全省大数据应用技能大赛（以下简称“大赛”），现将有关事项通知如下：

一、赛事组织

大赛由省总工会、省人力资源和社会保障厅、省营商环境建设局（省大数据局）、省委网络安全和信息化委员会办公室、省工业和信息化厅联合主办，辽宁省辽商总会承办。成立大赛组委会及其工作机构，负责大赛的各项组

织和协调工作。大赛组委会由主办、承办单位相关责任人构成，负责大赛组织管理和统筹协调等工作。

大赛组委会下设办公室。办公室设在辽宁省辽商总会软件和电子信息委员会，由赛事相关人员组成专家组、裁判组、赛务组、会务组、仲裁组、监审组，分别负责大赛组织、统筹协调、会务保障、裁判考核、监审仲裁和大赛活动总结、宣传推广等日常工作。

二、比赛方式

大赛为个人赛，由理论知识考核和实操考核两部分组成，按照《国家职业技能标准》有关规定要求实施，理论知识考核和实操考核试题由大赛专家组命制，理论知识考核采取闭卷上机考试；实操考核采取现场操作方式进行，成绩权重设定为理论知识考核（30%）实操考核（70%）。相关实施文件详见附件。

三、组织实施

（一）赛事安排

大赛拟于8-9月举行，具体时间地点另行通知（详见公众号“辽宁省软件行业协会官微”）。

（二）报名要求

大赛面向省内政府机关和企事业单位职工或高校教师，大赛参赛选手需具有2年以上从业经验，各大赛报名要求详见“辽宁省软件行业协会官微”（不收取报名费用），报名选手需经审核通过后方可参赛。

（三）人员要求

参赛选手须为从事或热爱大数据相关领域的在职职工，往届大赛荣获“辽宁五一劳动奖章”和“辽宁省技术能手”及以上荣誉的不得以选手身份参赛。

（四）运行方式

大赛分为初赛、决赛两个阶段；

初赛阶段。通过大赛报名平台成功报名的参赛选手，成绩前 50 名的入围决赛，初赛成绩不带入决赛。

决赛阶段。决赛由大赛组委会按照实施方案组织实施，决赛中选手理论知识考核和实操考核成绩按照权重累加得分，进行综合排名。

四、奖励办法

1. 对荣获大赛决赛第 1 名的选手，符合条件的，原则上推荐授予“辽宁五一劳动奖章”和“辽宁省技术能手”称号；给予 20000 元资金奖励并颁发名次证书。

2. 对荣获大赛决赛第 2 至 5 名的选手，由大赛主办单位联合颁发名次证书，依次给予 15000 元、10000 元、8000 元、5000 元资金奖励。

3. 对入围大赛决赛的选手，由大赛组委会颁发优秀选手证书，给予 1000 元奖金奖励。奖金和证书与决赛荣获名次选手不兼得。

五、相关要求

1. 大赛有关单位要坚持全省一盘棋，做好相互配合共同推进，积极组织职工参赛。承办单位要高度重视，做好谋划设计，广泛发动，注重细节，精心组织，系统推进，按时间节点把各项工作落到实处。

2. 坚持培训和大赛一体化推进，既要注重选拔高技能人才，又要注重职工参与率和有效覆盖面，推动职工队伍职业技能素质得到整体提升。

3. 坚持公平公正，严格按照大赛规则组织开展大赛活动，对暗箱操作、以权谋私等违纪违规行为，按有关规定严肃处理。

4. 保证大赛经费运行安全，严格按照有关财务规定管理使用大赛经费，对违反财经纪律的行为，按照有关规定追究当事者责任。

5. 做好大赛宣传，充分利用广播、电视、报刊、网络等多种媒体，对大赛中涌现出的优秀选手进行全过程宣传报道。

6. 加强大赛期间疫情防控工作，承办单位要与当地卫生健康部门沟通联系，严格落实疫情防控各项措施，科学制定竞赛期间疫情防控工作方案和应急处置预案，做好疫情防控工作，确保赛事活动稳妥、安全、有序开展。

六、联系方式

大赛组委会办公室：辽宁省辽商总会软件和电子信息委员会

大赛咨询：张秀丽 024-23447421、18249783860

大赛报名：关注“辽宁省软件行业协会”官方微信公众
众号，实名制线上统一报名入口。

附件：2022 辽宁省职工技能大赛暨第二届全省大数据
应用技能大赛实施方案

辽宁省职工技能大赛工作领导小组办公室

2022年8月15日



附件 1

2022 辽宁省职工技能大赛暨第二届全省 大数据应用技能大赛实施方案

一、竞赛目的

2022 辽宁省职工技能大赛暨第二届全省大数据应用技能大赛是贯彻落实国务院发布的《促进大数据发展行动纲要》，加快实施国家大数据战略，推动大数据产业健康快速发展，并紧密结合我省大数据产业发展战略规划和大数据技术发展方向，贯彻省委省政府《数字辽宁、智造强省》中人才发展战略，全面考察选手在企业真实项目环境下进行大数据平台部署管理、Hadoop 及其生态组件部署与管理、大数据技术基础、软件开发相关技术、数据采集与预处理、数据清洗与挖掘分析、数据可视化及综合分析的能力。

二、报名要求

大赛为个人赛，参赛选手需具有 2 年以上从业经验，各大赛报名要求详见“辽宁省软件行业协会官微”（不收取报名费用），报名选手需经审核通过后方可参赛。往届大赛荣获“辽宁五一劳动奖章”和“辽宁省技术能手”及以上荣誉的不得以选手身份参赛。

三、竞赛形式

本赛项为个人赛，共分为初赛和决赛两个阶段。

1. 初赛阶段

拟采用线上形式，分为理论知识考核和实操考核，比赛时间 2 小时，理论知识考核 0.5 小时、实操考核 1.5 小时。理论知识考核 100 分（单选、判断总计 50 道题），实操考核 100 分，成绩权重为理论知识考核 30%和实操考核 70%，成绩累加综合分数前 50 名选手进入决赛阶段。

2. 决赛阶段

采用线下形式，分为理论知识考核 0.5 小时+实操考核 5.5 小时，总时长 6 个小时（比赛期间不设午休，统一提供午餐和小食，竞赛不暂停，选手自主安排午餐时间，但不可自由走动）。理论知识考核 100 分（单选、判断总计 50 道题），实操考核 100 分，成绩权重为理论知识考核 30%和实操考核 70%。

本赛项为个人赛。竞赛结果评判严格按照大赛组委会的指导规范要求，借鉴世界技能大赛大数据赛项竞赛模式，由竞赛系统全自动判分，实时展示得分、排名。比赛结束后由专家对评分结果进行抽检复核，严格体现竞赛的公平、公正、公开的原则。

四、竞赛时间、地点

大赛拟于 8-9 月举行，具体时间地点另行通知（详见公众号“辽宁省软件行业协会官微”）。

五、竞赛内容

（一）理论知识考核

本赛项理论知识考核内容主要涉及大数据平台部署管理、数据采集与处理、数据清洗与挖掘分析、大数据管理及综合分析等方面知识点。

（二）实操考核

该赛项以大数据中较典型的互联网数据应用为技能实践的比赛内容，通过给定的任务需求，在规定的比赛时间内完成任务要求，进行技能实操考核。考核内容主要包括大数据平台及组件的部署管理、集群管理与运维、大数据处理与应用、大数据分析挖掘（Hadoop）、大数据分析挖掘（spark）、算法考核、数据可视化。

1. Hadoop、Spark、zookeeper、Hive 平台的配置与部署，考察 Hadoop、Spark、Hive 集群的部署能力，掌握常用系统配置和基础软件安装，能够部署和管理集群。

2. 集群管理与运维。考察大数据平台参数调整和实际运用能力，掌握集群维护、增加节点，能够根据程序异常信息进行分析和参数调优。

3. 大数据处理与应用。掌握 Linux 基础命令、HDFS Shell 操作，Hive SQL 命令，熟悉数据仓库架构，对数据进行分层处理。

4. 大数据分析挖掘（Hadoop）。掌握 Java 编程，Hadoop API 的使用。考察 Flume 日志采集、Hadoop 文件系统的常用命令，掌握分布式文件系统命令行操作与 API 使

用。掌握协同过滤推荐算法的原理，掌握评分矩阵、同现矩阵、合并矩阵、推荐向量的代码实现。

5. 大数据分析 with 挖掘 (spark)。考核对分布式计算和分布式存储架构下的数据读写、数据清洗、数据转换、数据分析等综合应用能力，使用 Scala 语言完成 Spark 框架下的 HDFS 读写，函数关联、数据切分、数据分析等数据操作。

6. 算法考核。考核选手对于数据各数值变量的操作，对于异常值、缺失值、重复数据的操作。使用 RFM 模型对数据进行分组，考核选手按照等深分箱的方法计算 R、F、M 的得分以及 RFM 总得分，查看数据分布是否均匀，对于制定特征变量进行分析和建议。

7. 数据可视化。通过常见的数据可视化方法，将数据分析结果以图表形式进行呈现，使用 Echarts，实现数据源分析结果展示。

六、竞赛环境

类别	名称	备注
硬件资源	路由交换模块	通用网络设备，三层交换机，千兆 RJ45 接口 16 个以上。
	PC 机	i5 以上 CPU 或同性能其他 CPU，内存 8G 以上，SSD 硬盘 128G 以上。
软件资源	云资源	独立大数据平台租户，资源配额不小于 16 个 vCPUS、32G 内存。
	竞赛平台	大数据竞赛系统，云平台。

七、成绩评定

（一）排名规则

1. 大赛不设并列奖项，依据理论知识考核、实操考核成绩之和进行排名，即总成绩较高者名次在前；
2. 总成绩相同者，实操考核成绩高者名次在前；
3. 实操考核成绩相同者，实操考核比赛所用时间少者名次在前；
4. 如以上完全相同者，由裁判组根据选手实际比赛情况确定名次，大赛不予以加赛处理。

（二）组织分工

1. 本竞赛参与赛项成绩管理的组织机构包括裁判组、监督组和仲裁组。
2. 监督组对裁判组的工作进行全程监督，并对竞赛成绩抽检复核。
3. 仲裁组负责接受由参赛队领队提出的对裁判结果的申诉，组织复议并及时反馈复议结果。
4. 竞赛将制定裁判遴选管理办法、赛事保密细则和预案、命题管理办法等制度，保证竞赛的公平公正。

（三）评分方法

1. 本赛项采用竞赛系统评分和结果评分。每名选手一个账号和密码，竞赛结束前保存成果并提交。
2. 裁判组汇总所有选手的评分表，计算成绩，裁判长及裁判成员签字确认，成绩汇总表备案以供核查。
3. 为保障成绩评判的准确性，监督组将对赛项总成绩

排名前 20 的所有参赛队伍的成绩进行复核；对其余成绩进行抽检复核，抽检覆盖率不得低于 15%。如发现成绩错误以书面方式及时告知裁判长，由裁判长更正成绩并签字确认。复核、抽检错误率超过 5%的，裁判组将对所有成绩进行复核。

4. 竞赛成绩经复核无误后，经裁判长、监督人员审核签字后公布。

八、竞赛细则

（一）选手须知

1. 参赛选手应遵守考场纪律，服从监考人员管理。违者取消本次竞赛成绩。

2. 参赛选手必须按竞赛时间，提前 30 分钟到达赛场。迟到 15 分钟者不得参加竞赛。

3. 参赛人员应凭本人身份证进入考场。

4. 决赛比赛环境为有线网络，大赛组委会统一提供电脑及相应软件环境。

5. 赛场不准带入技术资料 and 任何工具书，所有存储设备（U 盘）、通讯工具一律不得带入竞赛现场。

6. 选手在竞赛过程中不得擅自离开赛场，如有特殊情况，需经监考人员同意后作特殊处理。

7. 参赛选手在竞赛过程中，如遇问题需举手向监考人员提问，选手之间互相询问按作弊行为处理。

8. 在竞赛规定时间结束时应立即停止答题或操作，不

不得以任何理由拖延竞赛时间。

9. 待赛人员，应在指定地点等候（候赛室），听从指挥，不得擅自出入。

（二）工作人员须知

1. 树立服务观念，一切为选手着想，以高度负责的精神、严肃认真的态度和严谨细致的作风，积极完成本职工作。

2. 注意文明执裁，保持良好形象，熟悉竞赛方案。

3. 于赛前 45 分钟到达赛场，严守工作岗位，不迟到，不早退，不脱岗。

4. 熟悉竞赛规程，严格按照工作程序和有关规定办事，遇突发事件，按照安全工作预案，组织指挥实施，确保人员安全。

5. 服从统一领导，严格遵守竞赛纪律，加强协作配合，提高工作效率。

6. 如遇突发事件要及时报告，同时做好疏导工作，避免重大事故发生，确保大赛圆满成功。

（三）比赛规则

1. 大赛相关的各种设备由大赛组委会指定。

2. 参赛选手在比赛前进行抽签来决定比赛座位工位。

3. 比赛前 30 分钟进入比赛工位，比赛开始前 10 分钟，讲解考试注意事项，以场地计时器为准计时进行比赛（考场准备计时器固定在明显位置）。

4. 每场比赛连续进行，比赛过程中，食品和饮用水由赛场统一提供，选手休息、饮食或如厕时间都计算在比赛时间内。

5. 比赛期间参赛选手不得离场，不得携带手机、无线上网卡、移动存储设备、资料等与竞赛无关的物品。

6. 比赛过程中，参赛选手须严格遵守安全操作规程及劳动保护要求，确保设备及人身安全，并接受裁判员的监督和警示。

7. 因设备自身故障导致选手中断竞赛，由大赛裁判长视具体情况做出裁决。

8. 参赛选手若提前结束比赛，应向裁判员举手示意，比赛终止时间由裁判员记录，并清理现场卫生，参赛选手结束比赛后不得再进行任何操作。

9. 在比赛进行期间，竞赛场地内将开启信号干扰器、信号屏蔽器等设备，屏蔽现场的手机信号和 WLAN 信号等。

（四）申诉仲裁

1. 参赛选手对不符合竞赛规定的设备、工具和备件，以及对工作人员的违规行为等，均可提出申诉。

2. 选手申诉须由本人在所参加竞赛结束 2 小时内用书面形式向仲裁工作组提出。

3. 为保证竞赛顺利进行，保证竞赛结果公平公正，大赛组委会下设仲裁工作组，仲裁工作组负责受理大赛中出现的所有申诉并进行仲裁。

4. 仲裁工作组要认真负责地受理选手申诉，并将处理意见尽快通知领队或当事人。

5. 仲裁工作组的裁决为最终裁决，参赛选手不得因申诉或对处理意见不服而停止竞赛，否则按弃权处理。

九、题目案例

例题 1	
考核大项	考核内容
基础搭建	更改本地源
	配置 hosts 文件
	关闭防火墙
	时间同步
	下载 ntp
	制作定时任务
	配置 ssh 免密
安装 JDK	安装 JDK
安装 zookeeper	下载软件
	配置 zookeeper 环境变量
	配置文件 zoo.cfg
	启动 Zookeeper 集群
安装 Hadoop	解压安装包。配置环境变量
	配置 Hadoop 各组件
安装 hive	slave2 上安装 mysql server
	数据仓库下载安装
	hive 元数据配置到 MySQL (服务器端 slavel)
	Master 作为客户端配置 hive
成功启动 Hive	成功启动 Hive
	贷款数据分析 (在线借贷用户画像分析: 基于大数据的消费信贷算法制作贷款违约群体画像, 实现客户信息风险评估, 为贷款审核过程提供违约预警提示)
集群节点增加与删除	添加 DataNode-----slavel3
	删除 DataNode-----slavel2

例题 2

考核大项	考核分项	考核内容
构建数据仓库	环境准备	修改 hosts 文件
		确认防火墙为关闭状态
		启动 zookeeper 服务
		启动 Hadoop 集群
	master 下载安装 MySQLserver	安装 mysql-community-server
		开启 MYSQL 服务
		进入 MYSQL, 设置密码安全策略。密码强度为 low, 密码长度为 4
		修改 MSL 的本地 root 用户密码为 123456
		使用新密码重新登录 mysql, 为 root 用户设置远程连接权限, 并刷新权限
		在 MYSQL 中建立一个数据库 hongya
	Hive 基础环境配置	将提供好的 Hive 安装包解压到指定路径/usr/hive
		修改/etc//profile 文件设置 hive 环境变量
		修改 hive-site.xml 文件, 设置元数据位置
		修改 hive-site.xml 文件, 使用本地服务连接 Hive
		修改 hive-site.xml 文件, 配置连接 Mys
		置 ss 连接方式为 false
		设置 hive-site.xml 文件, 设置 MYSQL 驱动
		修改 hive-site.xml 文件, 设置连接用户名
		修改 hive-site.xml 文件, 设置连接密码
		修改 hive-site.xml 文件, 关闭 metastorel 的 schema 一致性
修改 hive-env.sh 文件, 配置 Hive 中 conf 文件路径		
将 MYSQL 的 JDBC 驱动包放至 hive 对应目录下		
更改 Hadoop 中 jine 的版本避免版本冲突		
成功启动 Hive		
安装 scoop	安装 scoop	将提供好的 scoop 安装包解压到指定路径
		修改/etc/profile 文件设置 scoop 环境变量
		将 MYSQL 驱动复制到 scoop 对应目录下
		修改 scoop-env.sh 文件, 设置 Hadoop 路径
		修改 scoop-env.sh 文件, 设置 zookeeper 路径
		修改 scoop-env.sh 文件, 设置 hive 路径
		scoop 中读取 hive 配置
		查看 scoop 版本
		列出 MYSQL 数据库列表, 验证 scoop 安装成功
数据分析	导入数据	hive 库下创建 data 表, 要求字段包括 tid, author, score, grade, titlecontent
		将数据存放至指定的 hive 库中的 data 表

	统计总用户数	统计总用户数,并将最后数目写入 /root/college022/目录下
	统计活跃用户	统计活跃用户 top10,将结果用户名及对应发帖数目 写入/root/college023/目录下
	查看用户等级	查找结果 4 中排序 no. 1 的用户等级,结果写入 /root/college024/(200)
	查看积分排行	根据积分排名,找出 top5 的用户,结果写入 /root/college025/目录下
单车数据 分析及可 可视化	单车数据预处理	将给出的数据集上传到 ndfs 的/college/目录下
		创建数据库 hive
		在 hive 数据库下构建数据表 bike
		将提供的分析数据导入到表 bike 中,并统计数据
	统计本次数据所有共享单车数量	统计本次数据所有单车数量,结果写入本地 /root/college001/中。
	单车平均用时计算	计算单车平均用时,结果写入本地 /root/college002/中。
	常年用车紧张的地区站点统计	统计常年用车紧张的地区站点,结果写入本地 /root/college003/中。(以 stratstation 为准)
	国庆期间各地区车辆使用情况统计	统计国庆期间各地区车辆使用情况,结果写入本地 /root/college004/中。(以 stratstation、 startdate 为准)
	共享单车单日租赁排行	给出共享单车单日租赁排行榜,结果写入本地 /root/college005/中。(以 startstation 为准)
	给出单车建议维修名单	给出建议维修的单车编号,结果写入本地中 /root/college006/。
	可进行会员活动推广的地区	给出可进行会员活动推广的地区,结果写入本地 /root/college007/中。(以 stratstation 为准)
	可舍弃的单车站点推荐	给出可舍弃的单车站点,结果写入本地 /root/collegeof08/中。(以 endstation 为准)
	根据数据挖掘计算开设共享大巴的可行性(1000)	根据 apron 算法,将支持度写到本地/root/college 中(保留五位小数)
		根据 apriori 算法,将置信度写到本地 /root/college010 中(保留五位小数)
上半年不同用户类型之间的用车情况示意图	制作上半年不同用户类型之间的用车情况示意图	
12 月份用车紧张地区示意图	绘出 12 月份用车紧张地区示意图	
人口数据 分析及可 可视化	人口数据预处理	将给出的数据 person.csv 上传至 hdfs 的 /college/目录下
		检查是否存在数据库 hive
		在 hive 数据库下创建数据表 person
		将数据导入至表 person 中,并统计数据

较高收入人群占整体数据的比例计算	统计较高收入人群占整体数据的比例（保留两位小数）
本科学历人员在调查中的占比计算	计算学历为本科的人员在调查中的占比（保留两位小数）
青年群体中高收入年龄层分析	计算青年群体中高收入年龄层排行
男性群体中高收入职业分析	计算男性群体中高收入职业排行
未婚女性高收入职业分析	对未婚女性高收入职业排行
分析性别对于收入的影响	统计性别对于收入的影响
分析教育程度对于收入的影响	统计教育程度对于收入的影响
不同收入对每周工时的影响	计算不同收入的平均工作时间
根据数据挖掘计算资本收益和职业的关联关系	根据 apriori 算法, 将支持度写到本地 /root/college019/口中(保留五位小数)
	根据 apriori! 算法, 将支持度写到本地 /root/college020/口中(保留五位小数)
绘出高收入人群教育程度分布图	绘出高收入人群教育程度分布图
绘出不同工作类型中高收入人群占比柱状图	绘出不同工作类型中高收入人群占比柱状图