**2017年全国职业院校技能大赛高职组**

**“大数据技术与应用赛项”赛项规程**

# 一、赛项名称

赛项编号：GZ-2017023

赛项名称：大数据技术与应用

英语翻译：Big Data Technology And Application

赛项组别：高职组

赛项归属产业：电子信息产业---战略新兴产业

# 二、竞赛目的

通过本赛项的举办, 可以检验参赛选手掌握大数据业务分析方法和技术应用能力、运用Hadoop及相关工具软件解决具体业务问题的能力，展示职教改革成果及参赛师生良好精神风貌。

通过本赛项，可以搭建校企合作的平台，提升高职电子信息类及相关专业毕业生能力素质，满足企业用人需求，促进校企合作协同育人，对接产业发展，实现行业资源、企业资源与教学资源的有机融合，使高职院校在专业建设、课程建设、人才培养方案和人才培养模式等方面，跟踪社会发展的最新需要，缩小人才培养与行业需求差距，引领职业院校专业建设与课程改革。

# 三、竞赛内容

赛项以大数据为核心内容和工作基础，重点考察参赛选手在Hadoop平台环境下，对于大规模并行数据分析以及内存计算技术的应用能力。具体包括：

1.需要针对竞赛选题，设计构建业务逻辑模型，形成业务设计文档，考察选手的业务分析能力；

2.基于大数据实训平台，完成Hadoop平台环境部署与基本配置，综合运用数据抓取爬虫工具，数据抽取工具Sqoop/Flume/Kafka、分布式存储系统HDFS、分布式计算框架MapReduce/Yarn、内存计算Spark、流式计算Storm、数据挖掘方法库Mahout、数据仓库Hive、Python等开发语言工具和技术，匹配和连接数据源，完成大数据的采集，提取、清洗、转换、分析、挖掘操作，产生分析结果，并且实现可视化呈现；

3.依据项目应用需求和分析结果，完成标准化的大数据分析报告。

竞赛时间4小时，竞赛连续进行。

竞赛内容构成如下：

|  |  |  |
| --- | --- | --- |
| **考核任务** | **考核知识点和技能点** | **描述** |
| Hadoop平台安装部署 | Hadoop平台安装部署和基本配置 | 考察Hadoop平台环境理解和部署能力，掌握常用的基本配置和命令，能够管理和维护Hadoop集群。 |
| HDFS常用操作命令（查询文件类别、上传、删除文件、查询HDFS基本统计信息） |
| Hadoop集群节点的动态增加与删除 |
| Hadoop集群的负载均衡 |
| 数据抓取 | 规则文件数据和关系型数据库数据抓取以及数据同步（同Hadoop/Hive数据同步） | 考察学生对Hadoop平台下的各种应用类型数据的抓取和收集 |
| 日志类数据抓取 |
| 流数据抓取（了解并掌握Kafka） |
| 能够通过Python编程语言网络爬虫抓取工具，实现网络数据抓取 |
| 数据清洗、整理、计算和表达 | 基于MapReduce的Hive数据仓库的离线数据的数据整理和计算 | 考察学生对数据进行整理、计算，通过Pythonapi接口，对计算分析结果导出到Hdfs中，供报表，日志分析等需求使用等技能。  能够通过常用的数据挖掘算法结合数据仓库Hive应用实现数据的计算和表达。 |
| 基于Spark技术的批处理、实时流计算以及基于SQL语义的数据仓库 |
| 利用Mabout中的分类、聚类、频繁模式挖掘等数据挖掘算法，结合Hive数据库技术实现对数据的分析和展现。 |
| 综合分析 | 通过知识技能，实现对提供数据源的分析，通过Python实现数据的展现，并根据展现结果提出结论，然后对结论进行分析，作出分析报告。 | 考察学生对大数据技术与分析的综合操作能力和分析能力 |

竞赛各阶段分值权重和时间分布如下：

|  |  |  |
| --- | --- | --- |
| **序号** | **竞赛时间** | **分值权重** |
| Hadoop平台安装部署 | 100分钟 | 权重15% |
| 数据抓取 | 权重30% |
| 数据清洗、整理、计算和表达 | 140分钟 | 权重20% |
| 综合分析 | 权重30% |
| 团队分工明确合理、操作规范、文明竞赛 | | 权重5% |

# 四、竞赛方式

本赛项为团体赛，以院校为单位组队参赛，不得跨校组队。每支参赛队由3名选手（设场上队长1名）和不超过2名指导教师组成。

本赛项拟邀请境外代表队参赛。

# 五、竞赛流程

竞赛期间各项日程与流程安排如下表。

|  |  |  |
| --- | --- | --- |
| **日期** | **时 间** | **内 容** |
| 第一日 | 8:00-14:00 | 报到 |
| 15:00-16:00 | 选手熟悉赛场 |
| 16:00-16:30 | 领队会、赛前说明 |
| 领队抽取场次签及检录顺序号 |
| 第二日 | 8:00-8:30 | 开赛仪式 |
| 8:30-8:50 | 裁判长在监督长监督下，现场抽取竞赛技术参数 |
| 8:50-9:20 | 赛场检录；竞赛选手抽签、确定竞赛工位号 |
| 9:20-9:30 | 竞赛选手进入赛位，检查赛位设备及耗材；  现场裁判发放题目、宣布竞赛注意事项； |
| 9:30-13:30 | 竞赛选手完成竞赛任务 |
| 14:00-20:30 | 竞赛成绩评定，进行成绩汇总 |
| 第三日 | 8:00-9:00 | 公布竞赛成绩 |
| 9:00-10:00 | 闭赛与颁奖仪式 |

# 六、竞赛试题

（一）赛项执委会下设的命题专家组负责本赛项命题工作。

（二）本赛项为公开比赛赛题。在距国赛开始日前一个月，通过全国职业院校技能大赛指定的网络信息发布平台（http://www.chinaskills-jsw.org）公布竞赛赛题。但用于比赛的数据技术参数，在国赛开始前1小时，由裁判长在监督长监督下，在竞赛现场随机抽取1套可选技术参数，作为竞赛使用参数。

（三）公开比赛赛题共设计6项可选技术参数。对于可选技术参数的说明及范例，作为公开比赛赛题的组成部分同时发布。

（四）国赛开始前3日，按照公开赛题规定，由命题专家组编制10套可选数据技术参数。参数严格保密，不予公开。到达国赛现场后，存放在承办院校保密室中，保密室应全程监控。

（五）本赛项样题详见《附件一》。

# 七、竞赛规则

1.报名资格

参赛选手须为2017年度职院校全日制在籍学生，本科院校中高职类全日制在籍学生，五年制高职四、五年级在籍学生。

2.竞赛工位于竞赛当天采用抽签方式确定，由每队参赛队队长进行抽签并确认。

竞赛期间参赛选手不得离开竞赛工位。

3.竞赛所需的硬件设备、系统软件和辅助工具由执委会统一安排，参赛选手不得自带硬件设备、软件、移动存储、辅助工具、移动通信等进入竞赛现场。

4.参赛队自行决定选手分工、工作程序和时间安排。

5.参赛队在赛前10分钟进入竞赛工位并领取竞赛任务，竞赛正式开始后方可展开相关工作。

6.竞赛过程中，选手须严格遵守操作规程，确保人身及设备安全，并接受裁判员的监督和警示。如遇到设备或软件等故障，参赛选手应当举手示意；现场裁判应及时予以解决。如确因计算机软件或硬件故障，致使操作无法继续的，由现场裁判协同技术支持人员共同解决。

7.若因选手因素造成设备故障或损坏，无法继续竞赛的，由现场裁判报告裁判长。裁判长有权决定终止该队竞赛；若因非参赛选手个人因素造成设备故障，由裁判长视具体情况做出裁决。

8.竞赛结束后，参赛队要确认已成功提交所有竞赛文档，由参赛队队长签字确认，参赛队在确认后不得再进行任何操作。

9.现场竞赛结束，经现场裁判确认后方可离开赛场。

10.其它未尽事宜，将在赛前向各领队做详细说明。

# 八、竞赛环境

1.竞赛场地。竞赛场地分为：竞赛现场、裁判员休息区、指导老师休息区、开（闭）幕式会议区、服务区。其中，竞赛现场又划分为：检录区、场内竞赛区、裁判工作区、技术支持区。以上区域应保证良好的采光、照明和通风；应提供稳定的水、电和供电应急设备。

2.竞赛设备。场内竞赛区按照参赛队数量准备比赛所需的软硬件平台，为参赛队提供统一竞赛设备和备用设备。选手无需自带任何工具及附件。

3.竞赛工位。竞赛现场各个工作区配备单相220V/3A以上交流电源。每个比赛工位上标明编号；每个比赛间配有工作台，用于摆放计算机和其它调试设备工具等；配备3把工作椅。

4.技术支持区。为技术支持人员提供固定工位、电源保障。

5.服务区。提供医疗等服务保障。

6.赛场开放。竞赛环境依据竞赛需求设计，在竞赛不被干扰的前提下赛场面向媒体、行业专家开放；允许媒体、行业专家在规定的时段内沿指定路线进行现场参观。

7.竞赛场地应符合消防安全规定，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，现场临时用电满足《施工现场临时用电安全技术规范》JGJ46-2005的要求。

# 九、技术规范

竞赛项目基础标准（见表2）和软件开发标准（见表3）如下：

（一）基础标准

|  |  |
| --- | --- |
| **标 准** | **内 容** |
| GB/T 11457-2006 | 信息技术、软件工程术语 |
| GB8566-88 | 计算机软件开发规范 |
| GB/T 12991-2008 | 信息技术数据库语言SQL第1部分：框架 |
| 20120567-T-469 | 信息技术云数据存储和管理第1部分：总则 |
| 20120568-T-469 | 信息技术云数据存储和管理第2部分：基于对象的云存储应用接口 |
| GB/T 21025-2007 | XML使用指南 |
| GB/T 20009-2005 | 信息安全技术数据库管理系统安全评估准则 已发布 |
| GB/T 20273-2006 | 信息安全技术数据库管理系统安全技术要求 |
| 20100383-T-469 | 信息技术安全技术信息安全管理体系实施指南 |

（二）软件开发标准

|  |  |
| --- | --- |
| **标 准** | **内 容** |
| GB/T 8566 -2001 | 信息技术 软件生存周期过程 |
| GB/T 15853 -1995 | 软件支持环境 |
| GB/T 14079 -1993 | 软件维护指南 |
| GB/T 17544-1998 | 信息技术 软件包 质量要求和测试 |

# 十、技术平台

竞赛技术平台为新建技术平台，要求最少满足100支参赛队伍300人规模。

## （一）竞赛设备

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **设备类别** | | **数量** | **设备用途** | **基本配置** |
| 服务器 | 应用服务器 | 2台。  比赛使用1台，备份1台。 | 内嵌大数据实训管理系统，提供大数据实训、虚拟机操作服务。 | Intel E5-2630v3 2.4G 8.0QPI 20M 8C×2/256GB内存/2TB 7200转 SATA 6Gb 硬盘/支持RAID5/1Gb ×2管理网络接口/1Gb ×2数据网络接口/1+1冗余电源 |
| 数据服务器 | 15台。  采用集群管理方式。 | 支撑大数据实训管理系统运行使用。内嵌虚拟化资源管理控制端，作为虚拟化资源管理系统的计算资源、网络资源和存储资源的源节点。 | Intel E5-2630v3 2.4G 8.0QPI 20M 8C×2/256GB内存/2TB 7200转 SATA 6Gb 硬盘/支持RAID5/1Gb ×2管理网络接口/1Gb ×2数据网络接口/1+1冗余电源 |
| 客户端 | 每支参赛队伍3台。 | | 竞赛选手比赛使用。 | 2.0GHZ处理器，2G以上内存，显示器要求1024\*768以上 |
| 根据参赛团队数量，配备30%的备份机器。 | |

## （二）软件环境

|  |  |  |
| --- | --- | --- |
| **设备类别** | **软件类别** | **软件名称** |
| 服务器 | 操作系统 | Linux |
| Hadoop分布式计算平台 |  |
| Web站点 |  |
| 客户端 | 操作系统 | Windows7 |
| 开发语言 | Python2.7.13 |
| 数据分析工具 | Pandas 0.9.12 |
| Python集成开发环境 | pycharm-community-2016.3.2 |
| 浏览器 | Chrome |
| 文档编辑器 | Office 2007以上版本 |

## （三）竞赛技术平台

比赛技术平台：《北京四合天地大数据实训管理系统》。

《北京四合天地大数据实训管理系统》是一套基于JavaEE技术开发的实训系统。

系统从大数据环境搭建、数据采集、数据清洗、数据分析、报表生成等环境，贯穿大数据平台的技术知识点，内容包括Hadoop运维技术、HDFS、Hbase、Hive、MapReduce、Kafka、Spark、Storm、Mahout、JSOUP、MySQL、Echarts等，所涉及开发语言包括Java、Python、HTML、Javascript、CSS等。

学生可以“随时、随地”在一套系统中进行多种类型的大数据相关内容的训练及比赛，并且可以在线进行操作系统级的真机模拟训练测试、使用在线报表系统生成报表，充分体验和掌握大数据相关技能。

教师（裁判）可以方便组织和安排大数据实训（竞赛），并且可以实时了解学生的训练及测试状态。训练、测试完毕后，教师（裁判）可以查看每位学生的测评报告，并且了解各个学生对大数据相关知识点和操作技能的掌握情况。

系统以“实训”为功能设计基础，以“大数据”为内容设计基础，旨在提高学生的实训效率，解决传统大数据实训方式准备时间长、硬件和空间成本高、实训内容管理难的问题，可广泛应用在远程教学、远程培训、技能竞赛等领域。

# 十一、成绩评定

## （一）评分标准

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **一级指标** | **二级指标** | **考核知识点和技能点** | **描述** | **分值** |
| 客观试题 | Hadoop平台安装部署 | Hadoop平台安装部署和基本配置 | 考察Hadoop平台环境理解和部署能力，掌握常用的基本配置和命令，能够管理和维护Hadoop集群。 | 15 |
| HDFS常用操作命令（查询文件类别、上传、删除文件、查询HDFS基本统计信息） |
| Hadoop集群节点的动态增加与删除 |
| Hadoop集群的负载均衡 |
| 数据抓取 | 规则文件数据和关系型数据库数据抓取以及数据同步（同Hadoop/Hive数据同步） | 考察学生对Hadoop平台下的各种应用类型数据的抓取和收集 | 20 |
| 日志类数据抓取 |
| 流数据抓取（了解并掌握Kafka） |
| 主观试题 | 数据抓取 | 能够通过Python编程语言网络爬虫抓取工具，实现网络数据抓取 | （同上） | 10 |
| 数据清洗、整理、计算和表达 | 基于MapReduce的Hive数据仓库的离线数据的数据整理和计算 | 考察学生对数据进行整理、计算，通过Pythonapi接口，对计算分析结果导出到Hdfs中，供报表，日志分析等需求使用等技能。  能够通过常用的数据挖掘算法结合数据仓库Hive应用实现数据的计算和表达。 | 20 |
| 基于Spark技术的批处理、实时流计算以及基于SQL语义的数据仓库 |
| 利用Mabout中的分类、聚类、频繁模式挖掘等数据挖掘算法，结合Hive数据库技术实现对数据的分析和展现。 |
| 综合分析 | 通过知识技能，实现对提供数据源的分析，通过Python实现数据的展现，并根据展现结果提出结论，然后对结论进行分析，作出分析报告。 | 考察学生对大数据技术与分析的综合操作能力和分析能力 | 30 |
| 团队分工明确合理、操作规范、文明竞赛 | | | 5 |

## （二）评分方式

1.组织与分工

（1）参与大赛赛项成绩管理的组织机构包括：裁判组、监督组和仲裁组，受赛项执委会统一领导。

（2）裁判组实行“裁判长负责制”，设裁判长1名；加密裁判2名；现场裁判8名；评分裁判10名，共计21人。

（3）裁判及其他现场工作人员职责如下：

——裁判长：负责竞赛过程各类裁判的指定、更换与工作调整；负责竞赛技术参数的现场抽取；竞赛现场问题的快速处置与裁定；竞赛评分过程的组织与安排；竞赛成绩的审核与上报。

——加密裁判：负责组织参赛队伍（选手）抽签，对参赛队信息、抽签代码等进行加密、解密工作；

——现场裁判：按规定做好赛场记录，维护赛场纪律，评定参赛队的现场得分；

——评分裁判：负责对参赛队伍（选手）的比赛作品、比赛表现按赛项评分标准进行评定。客观性内容由2名裁判打分后按取平均分数的方式评定成绩，2名裁判所给分值差别超过该项成绩20%的，由裁判长另指定其他裁判评定成绩；主观性内容由5名裁判打分后按去掉最高、最低分后的平均成绩确定成绩。

——检录工作人员：负责对参赛队伍（选手）进行点名登记、身份核对等工作；

——监考工作人员：协助现场裁判做好负责工位范围内的秩序维持工作，不得在考场内随意走动；

——技术支持人员：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工位进行赛场技术支持保障。

（4）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。

（5）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。

2.成绩评定方法

成绩评定是根据竞赛考核目标、内容对参赛队或选手在竞赛过程中的表现和最终成果做出评价。本赛项的评分方法为结果评分，即：依据赛项评价标准，对参赛选手提交的竞赛成果和作答卷进行评价评分。

所有的评分表、成绩汇总表备案以供核查，最终的成绩由裁判长进行审核确认并上报赛项执委会。

3.成绩公布方法

赛项成绩在赛项结束后由赛项组委会负责公布最终成绩。任何组织和个人，不得擅自对大赛成绩进行涂改、伪造或用于欺诈等违法犯罪活动。

为保障成绩评判的准确性，监督组将对赛项总成绩排名前30%的所有参赛队伍（选手）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。

竞赛成绩经复核无误后，由项目裁判长、总裁判长、监督员审核签字后确定。在赛场及赛场外张贴纸质成绩。若有异议，经过规定程序仲裁后，按照仲裁结果公布比赛成绩。

# 十二、奖项设定

参照执行《2016年全国职业院校技能大赛奖惩办法》的有关规定，各赛项设参赛选手团体一、二、三等奖。以赛项实际参赛队总数为基数，一、二、三等奖获奖比例分别为10%、20%、30%（小数点后四舍五入）。

获得一等奖的参赛队伍的指导教师获“优秀指导教师奖”。

# 十三、赛项安全

赛事安全是大赛一切工作顺利开展的基础和前提，是赛项筹备和运行工作必须解决的核心问题。本赛项将按照《2016年全国职业院校技能大赛安全管理规定》要求，完成赛项的安全保障工作。

（一）比赛环境

1.赛项执委会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执委会要求排除安全隐患。

2.赛场周围要设立警戒线，防止无关人员进入发生意外事件。比赛现场内应参照相关职业岗位的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。

3.承办单位应提供保证应急预案实施的条件。对于断电、防火、防水、疾病等突发情况的处置，必须明确制度和预案，并配备急救人员与设施。

4.承办单位应制定开放赛场和体验区的人员疏导方案。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.大赛期间，承办单位须设置赛场安全管理岗位，建立安全管理日志。

6.参赛选手进入赛位、赛事裁判及相关工作人员进入工作场所，严禁携带通讯、照相摄录设备，禁止携带记录用具。如确有需要，由赛场统一配置、统一管理。

（二）生活条件

1.比赛期间，由执委会委托承办单位统一安排参赛选手和指导教师食宿。须根据国家相关的民族政策，安排好少数民族选手和教师的饮食起居。

2.比赛期间安排的住宿地应具有宾馆/住宿经营许可资质。以学校宿舍作为住宿地的，大赛期间的住宿、卫生、饮食安全等由执委会和承办单位共同负责。

3.大赛期间有组织的参观和观摩活动的交通安全由执委会负责。执委会和承办单位须保证比赛期间选手、指导教师和裁判员、工作人员的交通安全。

4.各赛项的安全管理，除了可以采取必要的安全隔离措施外，应严格遵守国家相关法律法规，保护个人隐私和人身自由。

（三）参赛队职责

1.各学校组织代表队时，须为所有参赛人员购买大赛期间的人身意外伤害保险。

2.各学校代表队组成后，须制定相关管理制度，并对所有选手、指导教师进行安全教育培训。

3.各参赛队伍须加强对参赛人员的安全管理，实现与赛场安全管理的对接。

（四）应急处理

1.比赛期间发生意外事故，发现者应第一时间报告赛项执委会，同时采取措施避免事态扩大。

2.赛项执委会应立即启动预案予以解决并报告赛区执委会。赛项出现重大安全问题可以停赛，是否停赛由赛区组委会决定。

3.事后，赛区执委会应向大赛执委会报告详细情况。

（五）处罚措施

1.因参赛队伍原因造成重大安全事故的，取消其获奖资格。

2.参赛队伍有发生重大安全事故隐患，经赛场工作人员提示、警告无效的，可取消其继续比赛的资格。

3.赛事工作人员违规的，按照相应的制度追究责任。情节恶劣并造成重大安全事故的，由司法机关追究相应法律责任。

# 十四、竞赛须知

## （一）参赛队须知

1.参赛队名称：统一使用规定的学校代表队名称，不得使用与所在省市、学校无关的名字作为队名。

2.参赛队组成：每支参赛队由3名参赛选手组成，须为同校在籍学生，其中队长1名。每支参赛队可配2名指导教师，指导教师须为本校专兼职教师。不接受跨省市、跨校组队报名；同一学校相同项目报名只能有1支参赛队。

3.各参赛院校应指定1名负责人任赛项领队，全权负责参赛事务的组织、协调和领导工作。

4.参赛选手及指导教师在报名获得确认后，原则上不再更换。如在筹备过程中，参赛选手和指导教师因故不能参赛，须由所在省级教育主管部门于赛项开赛10个工作日之前出具书面说明，经大赛执委会办公室核实后予以更换。允许队员缺席比赛；允许指导教师缺席比赛。

5.参赛队按照大赛赛程安排，凭赛项执委会颁发的参赛证和有效身份证件参加比赛及相关活动。

6.赛项执委会统一安排各参赛队在比赛前一天进入赛场熟悉环境和设施情况。

7.参赛队选手、领队和指导教师要有良好的职业道德，严格遵守比赛规则和比赛纪律，服从裁判，尊重裁判和赛场工作人员，自觉维护赛场秩序。

8.领队应负责赛事活动期间本队所有选手的人身及财产安全，如发现意外事故，应及时向赛项执委会报告。

9.各学校组织代表队时，须安排为参赛选手购买大赛期间的人身意外伤害保险。

## （二）领队、指导教师须知

1.严格遵守赛场的各项规定，服从裁判，文明竞赛。如发现弄虚作假者，取消参赛资格，名次无效。

2.领队和指导教师务必带好有效身份证件，在活动过程中佩戴“指导教师证”参加竞赛相关活动。

3.各代表队领队要坚决执行竞赛的各项规定，加强对参赛人员的管理，做好赛前准备工作，督促选手带好证件等竞赛相关材料。

4.在比赛期间要严格遵守比赛规则，不得私自接触裁判人员。

5.竞赛过程中，未经裁判许可，领队、指导教师及其他人员一律不得进入竞赛现场。

6.如对竞赛过程有疑议，由领队和指导教师负责以书面形式向大赛仲裁委员会反映，但不得影响竞赛进行。

7.对申诉的仲裁结果，领队要带头服从和执行，并做好选手工作。参赛选手不得因申诉或对处理意见不服而停止竞赛，否则以弃权处理。

8.领队和指导老师应及时查看大赛专用网页有关赛项的通知和内容，认真研究和掌握本赛项竞赛的规程、技术规范和赛场要求，指导选手做好赛前的一切技术准备和竞赛准备。

## （三）参赛选手须知

1.参赛选手应严格遵守赛场规章、操作规程和工艺准则，保证人身及设备安全，接受裁判员的监督和警示，文明竞赛。

2.参赛选手应按照规定时间抵达赛场，凭身份证、学生证，以及统一发放的参赛证，完成入场检录、抽签确定竞赛工位号，不得迟到早退。

3.参赛选手进入赛场前，须将身份证、学生证、参赛证交由检录人员统一保管，不得带入场内。

4.参赛选手凭竞赛工位号进入赛场，不允许携带任何电子设备及其他资料、用品。

5.参赛选手应在规定的时间段进入赛场，认真核对竞赛工位号，在指定位置就座。

6.参赛选手入场后，迅速确认竞赛设备状况，填写相关确认文件，并由参赛队长确认签字（竞赛工位号）。

7.参赛选手在收到开赛信号前不得启动操作。在竞赛过程中，确因计算机软件或硬件故障，致使操作无法继续的，经项目裁判长确认，予以启用备用计算机。

8.赛项任务书及相关资料，均保存在操作系统桌面的“大赛资料”中。参赛选手应在竞赛规定时间内完成任务书内容，将相应的文档等上传到平台指定位置。

9.参赛选手需及时保存工作记录。对于因各种原因造成的数据丢失，由参赛选手自行负责。

10.参赛队所提交的答卷采用竞赛工位号进行标识，不得出现地名、校名、姓名、参赛证编号等信息，否则取消竞赛成绩。

11.竞赛过程中，因严重操作失误或安全事故不能进行比赛的（例如因操作原因发生短路导致赛场断电的、造成设备不能正常工作的），现场裁判员有权中止该队比赛。

12.参赛期间，选手连续工作，餐饮等由赛场统一提供。选手休息、就餐和如厕时间均计算在比赛时间内。

13.在参赛期间，选手应注意保持工作环境及设备摆放符合生产操作规程。

14.在比赛中如遇非人为因素造成的设备故障，经裁判确认后，可向裁判长申请补足排除故障的时间。

15.参赛选手不得因各种原因提前结束比赛。如确因不可抗因素需要离开赛场的，须向现场裁判员举手示意，经裁判员许可并完成记录后，方可离开。

16.凡在竞赛期间内提前离开的选手，不得返回赛场。

17.竞赛操作结束后，参赛选手需要根据任务书要求，将相关成果文件拷贝至U盘，填写结束比赛相关确认文件，并由参赛队长签字确认（竞赛工位号）。因参赛选手未能按要求，将相应的文档等上传到平台指定位置及U盘的，竞赛成绩计为零分。

18.竞赛时间终了，选手应全体起立，结束操作。将资料和工具整齐摆放在操作平台上，经工作人员清点后可离开赛场，离开赛场时不得带走任何资料。

19.在竞赛期间，未经执委会批准，参赛选手不得接受其他单位和个人进行的与竞赛内容相关的采访。参赛选手不得将竞赛的相关信息私自公布。

20.符合下列情形之一的参赛选手，经裁判组裁定后中止其竞赛:

（1）不服从裁判员/监考员管理、扰乱赛场秩序、干扰其他参赛选手比赛，裁判员应提出警告，二次警告后无效，或情节特别严重，造成竞赛中止的，经裁判长确认，中止比赛，并取消竞赛资格和竞赛成绩。

（2）竞赛过程中，由于选手人为造成计算机、仪器设备及工具等严重损坏，负责赔偿其损失，并由裁判组裁定其竞赛结束与否、是否保留竞赛资格、是否累计其有效竞赛成绩。

（3）竞赛过程中，产生重大安全事故、或有产生重大安全事故隐患，经裁判员提示没有采取措施的，裁判员可暂停其竞赛，由裁判组裁定其竞赛结束，保留竞赛资格和有效竞赛成绩。

## （四）工作人员须知

1.竞赛现场设现场裁判组，裁判长1名，现场裁判若干名。裁判要秉公裁判，监督检查参赛队安全有序竞赛。如遇疑问或争议，须请示裁判长裁决，裁判长的决定为现场最终裁定。

2.赛场工作人员由赛项执委会统一聘用并进行工作分工，进入竞赛现场须佩戴执委会统一提供的吊牌。

3.赛场工作人员需服从执委会的管理，严格执行执委会制订的各项比赛规则，执行执委会的工作安排，积极维护好赛场秩序，坚守岗位，为赛场提供有序的服务。

4.赛场工作人员进入现场，不得携带任何通讯工具或与竞赛无关的物品。

5.参赛队进入赛场，现场裁判及赛场工作人员应按规定审查参赛选手带入赛场的物品，如发现不允许带入赛场的物品，交由参赛队随行人员保管，赛场不提供保管服务。

6.赛场工作人员在竞赛过程中不回答选手提出的任何有关比赛技术问题，如遇争议问题，应及时报告裁判长。

7.裁判长在赛前组织评分裁判，针对“评分细则”做统一培训。赛项合作企业应提供竞赛技术平台的互联网访问环境，以及必要的技术支持服务，帮助裁判长和专家组长熟悉系统，便于裁判长开展工作。

# 十五、申诉与仲裁

1.本赛项在比赛过程中若出现有失公正或有关人员违规等现象，参赛队领队可在比赛结束后2小时之内向仲裁组提出书面申诉。

2.书面申诉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，并由领队亲笔签名。非书面申诉不予受理。

3.赛项仲裁工作组在接到申诉报告后的2小时内组织复议，并及时将复议结果以书面形式告知申诉方。申诉方对复议结果仍有异议，可由省（市）领队向赛区仲裁委员会提出申诉。赛区仲裁委员会的仲裁结果为最终结果。

# 十六、竞赛观摩

竞赛通过以下方式供各参赛队观摩、交流。

1.设计观摩区，使用大屏幕实时显示竞赛现场实况；

2.在竞赛不被干扰的前提下部分开放部分赛场。观摩人员需佩戴观摩证件在工作人员带领下，在指定时间，沿指定路线、在指定区域内到现场观赛；

3.比赛作品展示：闭幕式展示优秀选手作品。领队及指导教师可在工作人员的引导下观摩、参观，但不得损坏、复制选手作品。

竞赛现场观摩应遵守如下纪律：

1.观摩人员需由赛项执委会批准，观摩的时间、距离、方式由赛项执委会结合赛项具体情况确定；

2.文明观赛，不得大声喧哗，服从赛场工作人员的指挥，杜绝各种违反赛场秩序的不文明行为；

3.观摩人员不得同参赛选手、裁判交流，不得传递信息，不得采录竞赛现场数据资料，不得影响比赛的正常进行；

4.对于各种违反赛场秩序的不文明行为，工作人员有权予以提醒、制止。

# 十七、竞赛直播

竞赛的全过程将通过摄录像进行记录。在赛前，针对赛题保密、设备安装调试、软件安装等关键环节进行实况摄录；竞赛现场从竞赛开幕式开始、到竞赛闭幕式结束，采用全程摄录方式，对比赛进行全过程实况转播；赛后，将制作优秀团队选手代表采访、优秀团队指导教师采访、裁判专家点评和企业人士采访等视频资料。

为促进各参赛队间的相互学习与交流，竞赛现场设直播设备。直播设备的设置不得影响选手比赛。

# 十八、资源转化

参照《2016年全国职业院校技能大赛赛项资源转化工作办法》，制定本赛项资源转化方案。

（一）资源推广

收集整理赛事举办过程形成的各种资料，包括：竞赛样题、试题库；参赛视频、优秀成果；评委、裁判、专家点评材料；优秀选手和指导教师访谈等，形成一套完整的技术资料，报送至大赛执委会，并通过网络方式分发给职业院校，以便各学校可以针对选手在比赛过程出现的问题进行针对性的训练和解决。

具体工作内容和工作要求列表如下：

|  |  |  |  |
| --- | --- | --- | --- |
| **内容** | **要求** | **完成时间** | **负责单位** |
| 风采展示 | 画面精美、伴音动听、播放流畅，时间长度15分钟左右的赛项宣传片及获奖代表优秀选手、指导教师风采展示片；竞赛过程音视频记录。 | 赛项闭赛后10日内 | 承办院校 |
| 点评材料 | 评委、裁判、专家点评 | 赛项闭赛后10日内 | 赛项执委会 |
| 访谈资料 | 优秀选手、指导教师访谈 | 赛项闭赛后30日内 | 赛项执委会 |
| 竞赛分析报告 | 竞赛试题、竞赛技能考核评分案例及技术分析报告 | 赛项闭赛后30日内 | 赛项执委会 |

（二）教学研讨

以赛项总结会、研讨会等形式，传播大赛的成功经验，扩大大赛成果的影响。组织行业专家、一线教师，结合大赛题目和选手作品，共同探索竞赛目标与人才培养目标、竞赛组织与教学模式改革、实作考核与教学考核方式、职业竞赛与职业素养养成的结合方法，在兼顾知识、技能、素质发展和项目过程系统化的原则下，探索技能竞赛项目和评价标准，与专业课程项目化教学过程的有效结合，实现赛项资源向专业教学资源的转化，进而影响本专业教学改革方向。

（三）师资培训

结合大赛竞赛内容和竞赛方式，以及行业技能要求、教育教学需求，邀请行业技能考核专家、高职教学能手、企业技术专家作为培训讲师，组织大数据技术与应用专业教师参加的专项师资培训2场。

（四）教材开发

以国家示范校、行业专家、企业专家为班底成立教材开发团队，以大数据技术与应用为切入点，结合技能竞赛资源，编制开发符合行业、企业发展规律，又能利于教学改革的，符合本专业应用方向的理实一体化课程及教材2本，经行业、企业专家审核后交出版社规划出版。

（五）校企合作

组织参赛院校领导及师生到企业参观座谈，积极引导企业与高职院校间的联系，架设起本专业人才培养的校企合作、校企直通的桥梁。以赛项组织过程为样板，充分利用竞赛专家和技术优势，帮助参赛院校将竞赛平台转化为教学实训平台，助其建立大数据技术与应用职业教育实践创新基地。

# 附件1：大数据技术与应用竞赛样题

## Hadoop平台安装部署（15分）

本部分考察知识技能点主要包括如下几个方面：

1.Hadoop平台安装部署和基本配置；

2.HDFS常用操作命令（查询文件类别、上传、删除文件、查询HDFS基本统计信息）；

3.Hadoop集群节点的动态增加与删除；

4.Hadoop集群的负载均衡；

## 数据抓取（30分）

数据抓取主要考察参赛选手对各种数据抓取方法的掌握和应用，主要包括以下方面：

1.规则文件数据和关系型数据库数据抓取以及数据同步（同Hadoop/Hive数据同步）；

2.日志类数据抓取；

3.流数据抓取（了解并掌握kafka）；

4.能够通过Python编程语言网络爬虫抓取工具，实现网络数据抓取；

## 数据清洗、整理、计算和表达（20分）

数据清洗、整理、计算和表达，是对参赛选手掌握抓取数据的整理、计算和分析，考察知识技能点主要包括以下方面：

1.基于MapReduce的Hive数据仓库离线数据的数据整理和计算；

2.基于Apache Spark技术的批处理、实时流计算以及基于SQL语义的数据仓库；

3.利用Mabout中的分类、聚类、频繁模式挖掘等数据挖掘算法，结合Hive数据库技术实现对数据的计算、分析和展现；

## 综合分析（30分）

通过上述知识技能，实现对提供数据源的分析，通过Python实现数据的展现，并根据展现结果提出结论，然后对结论进行分析，作出分析报告。

## 团队综合素质（5分）

团队分工明确合理、操作规范、文明竞赛。