**2017年浙江省高职高专院校学生技能大赛**

**“大数据技术与应用”竞赛暨全国大赛选拔赛赛项规程**

# 一、赛项名称

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

# 二、竞赛目的

本赛项的是为适应大数据应用产业对高素质技术技能型人才的职业需求。通过竞赛，可以使学生熟悉Hadoop平台操作技术、Hive分布式数据库建设方法、分布式计算框架MapReduce综合运用手段；可以培养学生大数据采集能力、大数据预处理能力、Java/Python应用编程能力、数据查询与分析能力、数据可视化展现能力、组织管理与团队协调创新能力。赛项的实施不仅能够检验高职大数据相关专业的教学水平、人才培养方式，更能激发学生对各类大数据技术的学习兴趣和求知渴望，提升学生多种职业素养和职业技能，发挥“以赛促教、以赛促学”，引领多层次、多类型大数据人才培养。

通过本次大赛，搭建校企合作的平台，推进专业设置与产业需求对接、[课程内容](http://www.baidu.com/s?wd=%E8%AF%BE%E7%A8%8B%E5%86%85%E5%AE%B9&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1Y3P1NhnyP-rHm3nHP9uHT10ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EnHbdrHnvn1mYnWn4rjfsnWDsn0)与职业标准对接、[教学过程](http://www.baidu.com/s?wd=%E6%95%99%E5%AD%A6%E8%BF%87%E7%A8%8B&tn=44039180_cpr&fenlei=mv6quAkxTZn0IZRqIHckPjm4nH00T1Y3P1NhnyP-rHm3nHP9uHT10ZwV5Hcvrjm3rH6sPfKWUMw85HfYnjn4nH6sgvPsT6KdThsqpZwYTjCEQLGCpyw9Uz4Bmy-bIi4WUvYETgN-TLwGUv3EnHbdrHnvn1mYnWn4rjfsnWDsn0)与工作过程对接，促进校企联合在具有行业特色大数据应用人才和大数据应用技术技能拔尖人才培养模式方面不断探索。同时，促进大数据行业资源、企业资源与教学资源相整合，引导企业和学校在大数据应用方面深化合作，融合发展。通过在竞赛期间举行的系列配套活动，扩大我省高职教育在全国的影响力和竞争力，在整体提升我省高职教育办学水平的基础上，为进一步打造高职名校、名专业提供动力。

# 三、竞赛时间与内容

（一）竞赛时间

本次竞赛时间安排为4小时，比赛日期为4月21日、22日两天，具体安排如下表所示：

|  |  |  |
| --- | --- | --- |
| **比赛日期** | **赛程时间** | **赛程任务安排** |
| 4月21日  （周五） | 13：00~14：40 | 选手报到 |
| 15：00~15：40 | 参赛选手一次抽签加密、熟悉竞赛场地 |
| 15：00~16：00 | 领队会议 |
| 4月22日  （周六） | 7：00~7：40 | 第一批次参赛队到侯考区报到，  选手二次抽签加密 |
| 7：40~8：00 | 第一批参赛队进入竞赛工位并领取竞赛任务书 |
| 8：00~12：00 | 第一批参赛队正式比赛 |
| 11：40~14：40 | 第二批参赛队到侯考区报到、二次抽签加密（竞赛工位号）及封闭 |
| 14：40~15：00 | 第二批参赛队进入竞赛工位并领取竞赛任务书 |
| 15：00~19：00 | 第二批参赛队正式比赛 |
| 19：00~24：00 | 裁判评分 |

说明：如协办企业提供比赛设备充足，则比赛将安排在一个批次内完成，正式比赛批次与时间安排以竞赛秩序册为准。

（二）竞赛内容

本次竞赛时间安排为4小时。竞赛内容及评判依标准依据大数据技术应用相关岗位的能力要求，坚持普及行业关键技术、强化岗位核心技能、重视职业道德素养和团队协作精神。选取电子商务或互联网类企业数据中心的典型应用场景，要求参赛选手充分理解典型业务应用环境，使用大数据应用相关技术和规定工具软件来解决实际业务问题。赛项通过Hadoop平台环境调试、大数据采集与预处理、分布式数据库建设与调优、数据查询分析与可视化展现等方面的重要知识点，来体现相关核心技能。重点考察高职学生面对具体大数据应用场景的业务分析和模型设计能力、分布式存储和计算综合运用能力、数据查询与分析能力。

参赛队员必须在规定时间内完成比赛内容并提交相关文件。

**本次竞赛主要知识点和技能点如下：**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **知识模块** | **考核技能点** | **分值** |
| 客观部分 | Hadoop配置 | 考察Hadoop平台基本配置和维护； | 15 |
| 大数据采集 | 规定时间内，抓取数据的维度、内容、大小，是否符合赛题要求；  数据抓取主要是对网络数据和本地数据文件等数据类型的抓取处理。 | 40 |
| 大数据预处理 | 数据的抽取、预处理、优化过程是否符合赛题要求；  以预处理算法的适用性,正确率，速度作为标准指标；  数据预处理主要包括数据清洗、整理、计算、表达等方式，数据预处理是对数据的理解和数据应用行业背景的分析和理解。 |
| 数据查询与分析 | 数据查询和分析算法的适用性、正确率、速度等作为标准指标 |
| 数据可视化展现 | 数据挖掘结果呈现易理解程度 |
| 可视化展示是否完成，展示结果是否正确  数据展现是基于数据分析的结果进行展现，通过数据展现能够通过数据得到可以应用结论。展现是对行业数据的理解和以及行业业务应用的理解。 |
| 主观部分 | 编程试题 | 根据试题要求完成数据抓取、分析、展现等应用，考察参赛选手的动手实操能力。 | 10 |
| 编写提交数据分析报告，按照文档规范编写文档，贴近业务需求等，通过分析提供的数据，并对数据进行分析，展现，提出结论，考察选手的综合分析、对应用行业的业务和数据的理解与分析能力。 | | 30 |
| 职业素养：团队分工明确合理、操作规范、文明竞赛 | | 5 |

# 四、竞赛方式

1.本赛项为团体赛，以院校为单位组队参赛，不得跨校组队，每个学校最多有1个队伍参赛。每支参赛队由3名参赛选手组成（设场上队长1名）。参赛选手在现场根据给定的竞赛任务书，在4个小时内，相互配合共同完成竞赛任务。

2.每支参赛队可配1～2名指导教师。但在比赛期间指导教师不允许进入竞赛场地进行现场指导或在赛场外通过通信设备进行远程指导。

3.报名截止后，各支参赛队不再允许临时更改由报名时确定的选手名单和指导教师名单，若在竞赛时发现有这种行为的组委会将直接取消其比赛资格。

4. 每个队伍每人配备一台PC，现场所有竞赛队伍共享使用位于服务器上的大数据实训管理系统。

5.竞赛统一规定所需的硬件、软件和辅助工具的规格，确保竞赛平台统一。

# 五、竞赛规则与须知

1. 竞赛规则与选手须知

1．本次竞赛的比赛时间为4个小时，参赛队分1～2批在一天内完成竞赛(具体视参赛报名的队伍总数而定)。参赛队的比赛批次和竞赛工位号通过抽签决定。比赛期间参赛选手不得离开比赛场地，如有特殊情况，需经裁判人员同意后方可离开，但离开期间的时间一律计算在比赛时间内。

2. 每批次参赛选手必须在规定时间内到赛场侯考区报到，报到时应持本人身份证和学生证，并携带（佩戴）大赛组委会签发的参赛证（胸牌）。在赛前10-20分钟进入竞赛工位并领取竞赛任务书，熟悉比赛考试题目、商定任务分工、核对材料（物品）清单，但不准答题和操作。

3．参赛队按照参赛时段进入比赛场地，自行决定选手分工、工作程序和时间安排，利用现场提供的所有条件，用尽可能短的时间完成竞赛任务，封场期间参赛选手不得离开封闭场所，如有特殊情况，需经现场工作人员同意后方可离开并及时回来进行封闭。

4．竞赛所需的硬件、软件和辅助工具统一提供，竞赛工位内部U盘等,参赛队不得使用自带的任何有存储功能的设备，如硬盘、光盘、U盘、手机、随身听等,大赛会为每个参赛队提供一个U盘,以供队伍内部三名选手在比赛期间材料拷贝与交流.离开赛场时，不得将与比赛有关的物品带离现场，包括比赛提供的U盘。

5．比赛过程中，参赛选手必须严格遵守赛场纪律，不得在赛场内大声喧哗，不得作弊或弄虚作假；同时，必须严格遵守操作规程，确保设备和人身安全，并接受裁判员的监督和警示。若因选手因素造成设备故障或损坏，无法进行比赛，裁判长有权终止该队比赛；若因非选手个人因素造成设备故障的，由裁判长视具体情况做出裁决，故障中断时间不计入比赛时间。

6．在比赛过程中，参赛选手如有疑问，应举手示意，考场裁判长应按要求及时予以答疑。如遇设备或软件等故障，参赛选手应举手示意，考场裁判长、技术人员等应及时予以解决。确因计算机软件或硬件故障，致使操作无法继续的，经考场裁判长确认，予以启用备用设备。

7．当听到比赛结束命令时，各参赛队应立即停止答题或操作，不得以任何理由拖延竞赛时间。参赛队欲提前结束比赛，应向现场裁判员举手示意并记录比赛终止时间，比赛终止后，不得再进行任何与比赛有关的操作。

8．竞赛操作结束时，各参赛队要按照大赛要求和赛题要求提交竞赛成果。同时在提交的竞赛成果及现场留下的资料中不得出现与本参赛队相关的所有可能信息（包括学校名称、代表地区、选手姓名、指导老师姓名、抽签号），否则该队竞赛成绩将被取消。

9．竞赛操作结束后，参赛队要确认是否已成功提交了竞赛成果，裁判员在比赛结果的规定位置做标记，并与参赛队长一起签字确认。

（二）参赛队须知

1.参赛队数量：每个院校限报1支参赛队伍。

2.参赛队组成：每支参赛队由3名同校2017年在籍学生组成，性别和年级不限,本赛项也暂不邀请境外代表队参赛。

3.指导教师组成：每支参赛队可以配置1～2名指导教师，指导教师须为本校在职在编人员并经报名资格审查后确定。指导教师一经确认后，就不允许更换，但允许指导教师在竞赛当天缺席。

4.每支参赛队伍的参赛选手在报名确认后，原则上不再允许更换。如在备赛过程中，选手因生病等特殊原因不能参赛的，由省教育厅审批同意后，按相关参赛选手资格补充或更换。

5.参赛队若对竞赛过程有异议，在规定的时间内由领队向赛项仲裁工作组提出书面报告。

6.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。

7.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。

8.参赛队领队应对本队参赛队员和指导教师的参赛期间安全负责，参赛学校须为参赛选手和指导教师购买意外保险。

（三）指导教师须知

1.严格遵守赛场的规章制度，尊重裁判，文明竞赛。不得在赛场外面进行张望、喧哗等。

2.比赛期间，指导教师不能进入现场进行指导，也不能在场外通过通信设备等手段远程指导学生进行比赛。

（四）评分人员须知

1.评分人员要认真阅读评分细则，严格执行评分标准，准确量分。

2.评分人员要严格按照评分细则的规定评分，及时、准确地将评分结果记录在相应的评分登记表中，并签名。

3.评分人员在评分过程中存在疑问时，应及时向项目裁判长咨询。

4.评分人员在评分过程中发现的试题问题，应及时向项目裁判长报告。

（五）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，在赛项执委会的领导下，按照各自职责分工和要求认真做好岗位工作。

2.所有工作人员必须佩带证件，忠于职守，秉公办理，保守秘密。

3.注意文明礼貌，保持良好形象，熟悉赛项指南。

4.自觉遵守赛项纪律和规则，服从调配和分工，确保竞赛工作的顺利进行。

5.提前30分钟到达赛场，严守工作岗位，不迟到，不早退，不无故离岗，特殊情况需向工作组组长请假。

6.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，按照应急预案，组织指挥人员疏散，确保人员安全。

7.工作人员在竞赛中若有舞弊行为，立即撤销其工作资格，并严肃处理。

8.保持通讯畅通，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。

# 六、成绩评定与奖励办法

（一）评分标准与方法

竞赛评分严格按照公平、公正、公开的原则，评分标准注重考察参赛选手大数据技术与应用各个方面的技术水平与能力：

1. 本竞赛设置裁判5人，包括裁判长1名，裁判4名。共分为2组，每组2名裁判。评分分初评和复议。
2. 初评时，由裁判长根据参赛队伍数量组织分配2组裁判人员同步对各参赛队的答题结果进行判定等级和打分，判级和打分原则根据的具体的现场评分细则进行。
3. 竞赛满分为100分。
4. 团队比赛的初评成绩=客观部分得分+主观部分得分。
5. 当同组2名裁判对赛题判定评分一致或2名裁判按前一节所述评分标准所做评分的等级有差异，但等级差异在1个等级内的，则认为该团队初评分数有效，初评分即为最终成绩；
6. 当同组2名裁判判定同一赛题的评分等级差异大于等于2个等级时，应报告裁判长复议。裁判长在全部初判完成后，组织4名裁判共同对需复议部分题目进行复评。
7. 复议阶段仅对上述第（6）条有争议部分赛题进行评分。
8. 需复议部分赛题的复评成绩计算方法：4名裁判的评分分别去掉最高分和最低分，按剩余2名裁判打分成绩的平均分计算。
9. 复议队伍的最终成绩=复评部分赛题成绩和+无争议部分初评成绩。

全部参赛队打分完毕后，裁判长确认各团队的成绩全部有效之后。则通过网站相关页面公布各参赛团队的成绩及其次名次。

（二）竞赛的奖励办法

1、竞赛只设团体奖，团体奖根据参赛代表队总得分，进行排序。一等奖获奖比例10%，二等奖获奖比例20%，三等奖获奖比例30%。

2、一、二、三等奖代表队将获得相应的获奖证书。

3、根据比赛成绩确定代表浙江省参加全国职业院校技能竞赛的队伍，具体名额将根据教育部的要求确定。

# 七、申诉与仲裁

（一）申诉

1.参赛队对不符合竞赛规定的设备、工具、软件，有失公正的评判、奖励，以及对工作人员的违规行为等可提出申诉。

2.申诉应在本赛项竞赛结束后1小时内提出，超时不予受理。申诉时，应按照规定的程序由参赛队领队向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。申诉报告须有申诉的参赛选手、领队签名。

3.赛项仲裁工作组在收到申诉报告后，应根据申诉事由进行审查，12小时内书面通知申诉方，告知申诉处理结果。如受理申诉，要通知申诉方举办听证会的时间和地点；如不受理申诉，要说明理由。

4.参赛队不得采取过激行为攻击工作人员，否则不予受理申诉；在约定时间内，如约定的联系人未到场或中途离开，视为放弃申诉。

（二）仲裁

大赛采取两级仲裁机制。赛项设仲裁工作组，赛区设仲裁委员会。

1.赛项仲裁工作组负责受理比赛中出现的申诉并进行仲裁，以保证比赛的顺利进行和比赛结果的公平、公正。

2.赛项仲裁工作组在接到申诉后的2小时内组织复议，并及时反馈复议结果。申诉方对复议结果仍有异议，可由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。

3.参赛队不得因提起申诉或对申诉处理意见不服而停止比赛或滋事，否则按弃权处理。比赛不因申诉事件而组织重赛。

# 八、竞赛技术规范

**1.竞赛项目行业、职业技术标准**

大数据技术与应用比赛过程中主要需要遵循有以下国家标准，参赛代表队在实施竞赛项目中要求遵循如下规范。

（1）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 |

（2）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 |

**2.竞赛场地和环境标准**

竞赛区每个工位配备3台PC，PC设备统一提供竞赛使用环境。

竞赛共享服务器资源，服务器提供大数据实训管理平台。

现场提供网络（不接入Internet）。

**3.竞赛技术平台**

（1）个人计算机，最低软硬件配置要求如下：

操作系统：Windows 7或更新版本

处理器：2GHz 32位或64位

内存：2GB（64位）或以上

硬盘：50GB（32位）/80GB（64位）或以上

显卡：支持DirectX 9 128MB或以上

显示器：分辨率1024x768像素或以上

（2）大数据实训管理系统包含：

竞赛使用的题库以及备选题库、考生资源。

管理系统提供对客观试题的自动评分功能。

（3）相关软件版本：

服务器端：

|  |  |  |
| --- | --- | --- |
| **序号** | **安装环境** | **介绍** |
| 1 | Linux7.3 | 操作系统 |
| 2 | 大数据实训平台 | 竞赛使用平台 |
| 3 | 网络数据资源 |  |

客户端：

|  |  |  |
| --- | --- | --- |
| **序号** | **软件** | **介绍** |
| 1 | Windows 7 | 操作系统 |
| 2 | Python2.7.13 | 开发语言 |
| 3 | Pandas 0.9.12，matplotlib，numpy，python-dateutil,pytz，xlwt，openpyxl | Python基本控件库 |
| 4 | pycharm-community-2016.3.2 | Python集成开发环境 |
| 5 | Chrome 56.02924.87 | 浏览器 |
| 6 | office 2013 | 文档编辑器 |

（4）辅助工具：

U盘（每个参赛队一个）