

2018 年安徽省高等职业院校技能大赛

大数据技术与应用赛项竞赛规程

一、赛项名称

赛项名称：大数据技术与应用

赛项组别：高职学生组

二、竞赛目的

本赛项的设置，是为适应大数据开发应用产业对高素质技术技能型人才的职业需求，围绕大数据处理相关算法应用与软件工具运用。赛项选取典型的大数据应用业务分析应用场景，重点检验参赛选手掌握大数据业务分析方法和数据分析能力、运用大数据平台及相关工具软件解决具体业务问题的能力，激发学生对大数据知识和技术的学习兴趣，提升学生职业素养和职业技能，努力为中国大数据应用产业发展储备及输送新鲜血液。

通过举办本赛项，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

赛项以大数据技术与应用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处理以及内存计算技术的应用能力。具体包括：

1. 掌握 Hadoop 平台环境部署与基本配置，了解基于大数据计算平台的常见应用；

2. 综合利用 numpy、pandas、matplotlib、scikit 模块和 MapReduce 技术、分布式存储系统 HDFS、分布式计算框架 MapReduce/Yarn、数据仓库

Hive、Python 等开发语言工具和技术，匹配和连接数据源，实现大数据的采集，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，完成数据分析报告的编写。

4. 竞赛时间 4 小时，竞赛连续进行。

竞赛内容构成如下：

考核环节	考核知识点和技能点
Hadoop 平台 安装部署	Hadoop 平台安装部署和基本配置。
	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 平台常见基础组件应用。
	Hadoop 集群的性能调优。
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）。
	掌握 Hive 数据仓库导入、导出以及同 Hadoop 数据文件的转换，掌握 Hive 数据管理常用命令。
	能够通过 Python 编程，使用 streaming 和 MapReduce 机制实现数据的抓取。
数据清洗、整 理、计算和表达	基于 MapReduce 的 HDFS 文件系统的文件数据的数据查询、整理和计算。
	应用 Python 编程工具，实现给定数据源的采集，提取、清洗、转换、分析、挖掘操作，产生分析结果，实现给定数据源的预处理。
	结合 Hadoop 大数据平台相关技术实现对数据的分析和展现。
综合分析	综合应用数据抓取、清洗、整理、计算和表达等相关知识及技术，使用 Python 编程工具，实现对提供的数据源分析、展现，根据展现结果得出结论，并对结论进行分析，形成分析

	报告。
--	-----

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	240 分钟	权重 20%
数据抓取		权重 25%
数据清洗、整理、计算和表达		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

大数据技术与应用赛项为团体赛。每支参赛队由 3 名参赛选手组成，须为同校在籍学生，不允许跨校组队，其中队长 1 名。

五、竞赛规则

（一）竞赛工位通过抽签决定。

（二）竞赛所需的硬件设备、系统软件和辅助工具由赛项组委会统一准备，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。

（三）严禁参赛选手、赛项裁判、工作人员私自携带通讯、摄录设备进入比赛场地。

（四）参赛队自行决定选手分工、工作程序和时间安排。

（五）所有参赛选手都必须携带参赛证件进行检录。参赛证件包含身份证、学生证、参赛证。

（六）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。参赛选手迟到超过竞赛时间一小时不得入场。

(七) 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。

(八) 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。裁判长有权决定中止该队竞赛; 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。

(九) 竞赛结束或提前完成后, 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(签工位号)确认, 参赛队在确认后不得再进行任何操作, 经现场裁判确认后方可离开赛场

(十) 成绩评定评分方法分为现场评分和结果评分, 经裁判长、监督人员签字确认后, 由大赛组委会公示后公布。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。

(十一) 比赛期间, 不安排指导教师进行现场指导和现场观摩。

(十二) 其它未尽事宜, 将在赛前向各领队做详细说明。

六、参赛对象

高职组参赛对象为全省高职院校 2018 年春季在籍学生(含本科院校高职学生, 1994 年 5 月 1 日之后出生)。

七、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、服务区。其中, 竞赛现场又划分为: 检录区、场内竞赛区、技术支持区。以上区域应保证良好的采光、照明和通风;

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。

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. 竞赛场地应符合消防安全规定，应急照明设施状态合格，赛场明显位置张贴紧急疏散图，赛场出入口专人负责。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。

八、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
-----------------	------------------

九、技术平台

(一) 竞赛设备

设备类别	数量	设备用途	基本配置
服务器	1 台。	内嵌大数据实训管理系统，提供大数据竞赛。	PC 服务器
客户端	每支参赛队伍 3 台。 根据参赛团队数量， 配备 10%的备份机 器。	竞赛选手比赛使用。	性能相当于 2.0GHZ 处理器， 4G 以上内存，显 示器要求 1024*768 以上

(二) 软件环境

设备类别	软件类别	软件名称
服务端	竞赛系统运行操作系统	CentOS
客户端	桌面操作系统	Windows7
	虚拟环境	Virtualbox
	虚拟操作系统	Ubuntu
	大数据平台软件	Hadoop
	开发语言	Python2.7
	数据分析工具	Pandas, Numpy, Matplotlib
	Python 集成开发环境	pycharm-community-2016.3.2
	浏览器	Chrome/firefox
文档编辑器	Office 2007 及以上版本	

(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》。

十、成绩评定

评分标准

考核环节	考核知识点和技能点	分值
Hadoop 平台 安装部署	Hadoop 平台安装部署和基本配置。	20
	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 平台常见基础组件应用。	
	Hadoop 集群的性能调优。	
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）。	25
	掌握 Hive 数据仓库导入、导出以及同 Hadoop 数据文件的转换，掌握 Hive 数据管理常用命令。	
	能够通过 Python 编程，使用 streaming 和 MapReduce 机制实现数据的抓取。	
数据清洗、 整理、计算 和表达	基于 MapReduce 的 HDFS 文件系统的文件数据的数据查询、整理和计算。	20
	应用 Python 编程工具，实现给定数据源的采集，提取、清洗、转换、分析、挖掘操作，产生分析结果，实现给定数据源的预处理。	
	结合 Hadoop 大数据平台相关技术实现对数据的分析和展现。	
综合分析	综合应用数据抓取、清洗、整理、计算和表达等相关知识及技术，使用 Python 编程工具，实现对提供的数据源分析、展现，根据展现结果得出结论，并对结论进行分析，形成分析报告。	30
	团队分工明确合理、操作规范、文明竞赛	5

(二) 评分方式

1. 竞赛成绩评定实行“裁判长负责制”，负责组织评分裁判进行成绩评定。

2. 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。

3. 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现场评分和结果评分，现场评分：现场裁判根据参赛队的操作规范以及综合表现情况进行评分；结果评分：依据赛项评价标准，对参赛选手提交的竞赛成果进行评分。

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组委会。

4. 成绩公布方法

竞赛成绩经复核无误后，经裁判长、监督人员审核签字后，以省教育厅最终公布结果为准。

十一、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，均可提出申诉。

2. 申诉应在竞赛结束后2小时内提出，超过时效不予受理。申诉时，应按照规定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将不予受理。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. 申诉人不得采取过激行为刁难、攻击工作人员，否则视为放弃申诉。

（二）仲裁

赛项设仲裁工作组接受由代表队领队提出的对裁判结果等方面问题的申诉。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。仲裁工作组的仲裁结果为最终结果。

十三、竞赛须知

（一）参赛队须知

1. 参赛队名称：每个院校不超过2个参赛队，统一使用规定的学校代表队名称，每个参赛队按照学校参赛队顺序编号；

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。

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，参赛选手因故不能参赛，须由所在上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，经大赛组委会办公室核实后予以更换。竞赛开始后，参赛队不得更换参赛选手，允许队员缺席比赛。不允许更换指导教师，允许指导教师缺席比赛现场。

5. 参赛队按照大赛赛程安排，凭大赛组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. 大赛组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，应及时向大赛组委会报告。

(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赛项的各项规定，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. 如对竞赛过程有疑义，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 and 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、学生证交由检录人员统一保管，不得带入场内。参赛证始终佩戴，以备检查。

4. 参赛选手凭竞赛工位号进入赛场，不允许携带任何书籍和其他纸质资料，竞赛统一提供草稿纸。不允许携带任何电子设备及通信工具和存储设备（如U盘），竞赛统一提供计算机以及应用软件。

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，赛场工作人员与参赛选手共同确认操作条件及设备状况，填写相关确认文件，并由参赛队长确认签字（签竞赛工位号）。

7. 参赛选手在收到开赛信息前不得启动操作。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，经裁判长确认，予以启用备用计算机。

8. 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。

9. 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。

10. 竞赛过程中，因严重操作失误或安全事故不能进行比赛的（例如因人为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。

11. 参赛期间，选手连续工作。选手休息、入厕等时间均计算在比赛时间内。

12. 在参赛期间，选手应注意保持工作环境及设备摆放符合企业生产“5S”（即整理、整顿、清扫、清洁和素养）的原则，如果过于脏乱，裁判长有权酌情处置。

13. 参赛期间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。

14. 参赛期间，选手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。

15. 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，并由现场裁判和参赛队长共同签字确认竞赛相关确认文件及U盘内容（签竞赛工位号）。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。

16. 参赛队提前结束比赛，应提交竞赛结果，向现场裁判举手示意，由裁判员记录竞赛终止时间，由现场裁判和参赛队长共同签字确认后可提前离开。

17. 在赛事期间，未经组委会批准，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。

18. 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，裁判要秉公裁判，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. 赛场工作人员由大赛组委会统一聘用并进行工作分工，进入竞赛现场须佩戴大赛组委会统一提供的胸牌。

3. 赛场工作人员需服从大赛组委会的管理，严格执行赛项各项比赛规则，

执行各项工作安排，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. 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。