**2019年焦作大学信息工程学院技能大赛**

**“大数据技术及应用”赛项竞赛规程**

一、项目名称

大数据技术及应用

二、竞赛目的

以焦作大学职业技能大赛为平台，通过举办本赛项，促进我校计算机专业建设，开展与大数据技术及应用相关的教育教学创新，促进实践教学内容、手段和方法的改革，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。

三、竞赛内容

赛项设计过程中，根据大数据企业的用人要求，分析人才所需知识技能，并据此设计出贴近真实工作环境的赛项内容。此外，赛项设置了大数据环境搭建、数据预处理、数据初步分析、数据分析展示等环节，通过竞赛项目考核锻炼选手如下能力：

1.大数据环境搭建与运维水平；

2.数据采集与预处理能力；

3.数据分析软件使用水平；

4.数据分析算法与挖掘能力。

具体竞赛内容如下：

|  |  |  |
| --- | --- | --- |
| **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） | 考察学生对Hadoop平台下的各种应用类型数据的抓取和收集 |
| 掌握Hive数据仓库导入，导出以及同Hadoop数据文件的转换，实现Hive数据管理常用命令。 |
| 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HDFS文件系统的文件数据的数据查询、整理和计算 | 考察学生对数据进行整理、计算，导出分析结果，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实现数据的计算和表达。 |
| 结合Hadoop大数据平台相关技术实现对数据的分析和展现。 |
| 应用Python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，实现可视化呈现。 |
| 综合分析 | 综合应用相关知识技能，实现对提供数据源的分析、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 |

四、竞赛方式

1.竞赛采取团体赛形式，参赛选手必须是本校在籍学生，每支参赛队由3名选手（设队长1名）和不超过2名指导教师组成（指导教师为本校专兼职教师）；

2.竞赛时间为2小时，采用小组合作的形式完成赛项任务，最后以提交的截图和文档作为最终评分依据；

3.不计参赛选手的个人成绩，以竞赛队的总成绩进行排序；

4.比赛过程中，全程封闭，非比赛工作人员不得进入赛场。

五、成绩评分

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 | 70分钟 | 权重15% |
| 数据抓取 | 权重30% |
| 数据清洗、整理、计算和表达 | 100分钟 | 权重20% |
| 综合分析 | 权重30% |
| 操作规范、文明竞赛 | | 权重5% |

五、比赛设备平台标准

1.硬件环境

计算机最低配置为AMD主频≥3200MHz或Intel主频≥2.0MHz

内存≥4GB；硬盘≥320GB；独立显卡显存512MB以上

2.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 |

六、奖项设定

大赛结束后，组委会在核准无误后向学校报送参赛选手的成绩。具体奖项由学校统一安排实施。本次竞赛拟按照参赛队数量的10%设立一等奖，20%设立二等奖，30%设立三等奖。

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，以及对工作人员的违规行为等，均可向竞赛裁判委员会提出申诉。

2.申诉应在竞赛结束后2小时内提出，超过时效将不予受理。申诉时，应按照规定的程序由参赛队领队向赛项裁判委员会递交书面申诉报告。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.竞赛裁判委员会收到申诉报告后，应根据申诉事由进行审查，6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.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。申诉人不满意赛项裁委会的处理结果的，可向大赛高职组赛事仲裁工作组提出复议申请。

（二）仲裁

1.裁判委员会设仲裁工作组，负责受理大赛中出现的申诉复议并进行仲裁，以保证竞赛的顺利进行和竞赛结果公平、公正。

2.仲裁工作组的裁决为最终裁决，参赛队不得因对仲裁处理意见不服而停止比赛或滋事，否则按弃权处理。

信息工程学院

二零一九年十月十日