

第二届“强国杯”技术技能大赛

——“信创大数据技术应用”赛项

技
术
方
案

2024年7月

目录

一、 大赛名称	1
二、 大赛意义	1
三、 大赛内容、形式和成绩计算.....	2
(一) 竞赛内容.....	2
(二) 竞赛形式.....	3
(三) 报名条件.....	3
(四) 成绩计算.....	3
四、 奖励办法	4
五、 大赛命题原则.....	5
六、 大赛范围、赛题类型和其他.....	6
(一) 大赛范围.....	6
(二) 赛题类型.....	6
七、 大赛场地与设施.....	7
(一) 大赛场地.....	7
(二) 竞赛设施.....	8
八、 大赛关键环节与时间安排.....	8

(一) 关键环节.....	8
(二) 竞赛流程.....	10
(三) 时间安排.....	11
九、 大赛评分标准制定原则、评分方法、评分细则及技术规范.....	12
(一) 评分标准制定原则.....	12
(二) 评分方法.....	13
(三) 评分细则(评分指标).....	13
(四) 技术规范.....	14
十、 大赛平台说明.....	15
(一) 竞赛设备.....	16
(二) 软件环境.....	16
十一、 大赛安全保障.....	17
(一) 组织机构.....	17
(二) 比赛环境.....	17
(三) 生活条件.....	18
(四) 参赛队职责.....	18
(五) 应急处理.....	19
十二、 大赛组织与管理.....	19

(一) 大赛组织.....	19
(二) 大赛设备与设施管理.....	19
十三、 附件	21

一、大赛名称

第二届“强国杯”技术技能大赛——信创大数据技术应用赛项。

二、大赛意义

习近平总书记指出“大数据是信息化发展的新阶段”，国家“十四五”规划也提出“打造数字经济新优势，充分发挥海量数据和丰富应用场景优势，促进数字技术与实体经济深度融合，赋能传统产业转型升级，催生新产业新业态新模式”，大数据成为推动社会发展的强大动力。

赛项以培养符合时代要求的高素质技术技能人才为使命，紧扣国家发展战略，通过竞赛形式，选拔和培养优秀的大数据技术人才，为我国信创产业和大数据领域提供源源不断的人才支持，从而推动信创产业与大数据技术的深度融合，提升我国在大数据领域的自主创新能力，加快实现信息技术自主可控。

大赛引领了专业建设和教学改革的方向，为学生提供了一个展示自我、锻炼能力的平台，推动职业教育与产业的深度互动。大赛的内容和标准紧密对接新技术、新产业、新业态、新模式，为职业教育改革提供了明确的指引。职业教育与产业的深度互动提高了职业教育的适应性，也为产业发展提供了有力的人才保障。

同时，大赛促进产教融合与科教融汇，优化了教育资源配置。学校与企业、科研机构的紧密合作，使得教育更加贴近产业需求，教学更加具有针对性。这种深度融合不仅提升了教育质量，也为产业发展注入了新的活力。

三、大赛内容、形式和成绩计算

(一) 竞赛内容

按照《大数据工程技术人员国家职业技术技能标准》要求，本竞赛项目结合行业、企业的实际业务场景和需求，培养和选拔“信创+大数据”人才，以容器云平台、国产数据库、麒麟操作系统为竞赛实训环境，涵盖大数据生命周期典型阶段技术，内容涉及大数据系统搭建和管理、数据采集、数据处理与分析、数据可视化、业务分析、职业素养。

1) 大数据平台搭建：重点体现国产操作系统（麒麟Linux、阿里龙蜥操作系统）应用与运维，基于容器云应用平台，搭建大数据系统环境，并与实际应用相结合，考核安装、配置和管理大数据相关组件，并发布应用程序等能力。安装Hadoop全分布式平台，安装Hadoop平台相关的常用组件，包括但不限于ZooKeeper、Flume、Kafka、Spark、Flink、Redis、HBase等，验证Hadoop平台和相关组件的可用性。

2) 数据库配置维护：根据预制场景，通过具有自主知识产权的数据处理与集成平台工具，采集给定数据源的数据，并对多源数据进行整合，熟练掌握有效数据存储、迁移和备份技术，确保数据全面性、多维性和高效性。基于国产化数据库进行建库建表，运用基本的SQL语句完成数据的增删改查等操作。

3) 数据获取与清洗：读取CSV数据源，对指定字段进行有效性检查，正确处理无效值和异常值，对数据进行一致性检查，对数据进行清洗和转换。

4) 数据统计：编写Java MapReduce程序，并将程序打包部署到Hadoop平台上运行，对数据进行统计汇总、分区分组和排序等操作。使用HDFS上传和下载文件。

5) 数据标注：使用Python程序对数据进行分类标注，将标注后的数据保存到指定位置。

6) 数据可视化：使用Web技术或Python可视化技术对数据进行呈现，包括但不限于柱状图、折线图、玫瑰图、气泡图、饼状图、条形图、雷达图、散点图等效果。

7) 业务分析：能够理解业务场景，对业务数据进行分析，编写分析报告。

8) 职业素养：本赛项为团体赛，团队分工明确合理、操作规范、文明竞赛。

(二) 竞赛形式

竞赛为线下比赛，组队方式为团体赛，以学校为单位组队报名参赛，每支参赛队伍需由3名参赛选手和不超过2名指导老师组成，每个参赛单位参赛队最多2支，同一参赛队成员须在同一单位，不得跨单位组队。

竞赛总时长为4小时，各参赛队在规定的时间内，独立完成“竞赛模块”规定的内容。

(三) 报名条件

参赛选手须为职业院校在校学生（包括中职、高职），该赛项以学校为单位组队报名参赛，每支参赛队伍需由3名参赛选手和不超过2名指导老师组成，每个参赛单位参赛队最多2支。同一参赛队成员须在同一单位，不得跨单位组队。

(四) 成绩计算

竞赛总时长为4小时，每个模块设置任务量4小时，各参赛队队员分

工完成。

模块		主要内容	任务时长	分值
模块一： 平台搭建 与运维	任务一： 大数据平台搭建	国产操作系统（麒麟Linux、阿里龙蜥操作系统）应用与运维，Hadoop 平台的安装部署和常用组件的安装部署。	120分钟	10
	任务二： 数据库配置维护	使用国产化数据库建库建表，运用基本的SQL语言完成数据的增删改查等操作。	120分钟	20
模块二： 数据获取 与处理	任务一： 数据获取与清洗	对CSV数据文件进行加载、清洗和转换等操作，识别和处理无效值，检查数据的一致性，将清洗后的数据保存到指定位置。	80分钟	10
	任务二： 数据标注	使用Python语言对数据进行分类标注。	80分钟	10
	任务三： 数据统计	基于Hadoop平台进行编译、打包、部署和执行程序，完成数据的统计工作。	80分钟	15
模块三： 业务分析 与可视化	任务一： 数据可视化	使用Web前端框架或者python可视化库对数据进行可视化展示。	120分钟	20
	任务二： 业务分析	报表分析，对大数据项目的业务场景和数据进行分析，撰写报告。	120分钟	10
职业素养		团队分工明确合理、操作规范、文明竞赛		5

四、奖励办法

各赛项均设置一二三等奖及优胜奖，颁发获奖证书，获奖比例原则上不超过10%、20%、30%及10%。对各赛项各组别一等奖获奖队伍的教练(每支参赛队伍指定1名教练)，颁发“优秀教练”证书；对做出突出贡献的承办单位和技术支持单位，颁发“突出贡献单位”证书；对大赛组织实施中表现突出的个人，颁发“优秀工作者”证书；对在各赛项执裁工作中表现突出的个人，颁发“优秀裁判员”证书。

五、大赛命题原则

按照《大数据工程技术人员国家职业技术技能标准》、相关 1+X 证书和《职业教育专业目录(2021年)》相关能力要求，本竞赛项目结合行业、企业的实际业务场景以及信创“开发、运维”核心岗位需求，培养和选拔“信创+大数据”人才，以容器云平台、国产化数据库、操作系统（麒麟 Linux）为竞赛实训环境，涵盖大数据生命周期典型阶段技术，内容涉及大数据系统搭建和管理、大数据采集、大数据处理与分析、大数据可视化、业务分析、职业素养，赛题遵循以下命题原则：

1 符合国家战略需求：命题紧密围绕国家信息技术创新发展战略，体现信创产业在大数据领域的前沿技术和应用趋势，以促进国家关键信息技术的创新发展。

2 行业实用性：命题注重实际应用场景，紧密贴合行业实际需求，使参赛者能够在实际工作环境中运用所学知识和技能，提高解决实际问题的能力。

3 技术创新性：命题鼓励参赛者运用新技术、新方法、新思路，突出创新性和创造性，激发参赛者的创新潜力，推动信创大数据技术的创新发展。

4 公平性和公正性：命题遵循公平、公正、公开的原则，确保竞赛的公平性和透明度，避免任何形式的不正当竞争和偏见。

5 可操作性和可评估性：命题确保竞赛任务的可操作性和可评估性，制定明确的评分标准，使评委能够准确、客观地评价参赛者的表现。

6 安全性和保密性：命题重视信息安全和保密要求，确保竞赛过程中不涉及敏感信息和核心机密，防范信息安全风险。

7 促进人才培养：命题着眼于促进信创大数据领域的人才培养，通过竞赛发现和选拔优秀人才，为信创产业的可持续发展提供人才保障。

六、大赛范围、赛题类型和其他

（一）大赛范围

本次“信创大数据技术应用”赛项旨在促进信创领域大数据技术的创新与应用，提高参赛选手在大数据采集、存储、处理、分析和应用等方面的综合能力。大赛将围绕信创产业的大数据技术发展趋势，结合实际应用场景，设置多个竞赛环节，全面考核参赛选手的技术实力和创新能力。

（二）赛题类型

实际操作竞赛：

（1）大数据平台搭建：重点体现国产操作系统（麒麟Linux、阿里龙蜥操作系统）应用与运维，基于容器云应用平台，搭建大数据系统环境，并与实际应用相结合，考核安装、配置和管理大数据相关组件，并发布应用程序等能力。安装Hadoop全分布式平台，安装Hadoop平台相关的常用组件，包括但不限于ZooKeeper、Flume、Kafka、Spark、Flink、Redis、HBase等，验证Hadoop平台和相关组件的可用性。

（2）数据库配置维护：根据预制场景，通过具有自主知识产权的数据处理与集成平台工具，采集给定数据源的数据，并对多源数据进行整合，熟练掌握有效数据存储、迁移和备份技术，确保数据全面性、多维性和高效性。基于国产化数据库进行建库建表，运用基本的SQL语句完成数据的增删改查等操作。

(3) 数据获取与清洗：读取CSV数据源，对指定字段进行有效性检查，正确处理无效值和异常值，对数据进行一致性检查，对数据进行清洗和转换。

(4) 数据统计：编写Java MapReduce程序，并将程序打包部署到Hadoop平台上运行，对数据进行统计汇总、分区分组和排序等操作。使用HDFS上传和下载文件。

(5) 数据标注：使用Python程序对数据进行分类标注，将标注后的数据保存到指定位置。

(6) 数据可视化：使用Web技术或Python可视化技术对数据进行呈现，包括但不限于柱状图、折线图、玫瑰图、气泡图、饼状图、条形图、雷达图、散点图等效果。

(7) 业务分析：能够理解业务场景，对业务数据进行分析，编写分析报告。

七、大赛场地与设施

(一) 大赛场地

赛场内设选手检录区、选手休息区、竞赛区、裁判组工作区、技术支持区、服务保障区、加密裁判封闭区、配件仓库、保密室等，

1. 选手检录区：选手等待检录、检录、一次加密、二次加密等职能区域；

2. 选手休息区：选手检录前及竞赛离场休息、指导教师休息区域；

3. 竞赛区：每个参赛队伍的比赛占地面积平均不少于 9 平方米。每个竞赛工位设工位编号，竞赛工位相对独立，确保选手独立开展竞赛，不受外界影响。赛场内安装摄像头，确保每个竞赛工位无盲区监控。赛

场环境的供电采用双强电设备，确保比赛用电的高可靠，各工位分区供电，强电弱电分开布线；场地采光、照明和通风良好，工位及竞赛桌面照度大于 500lux。

（二）竞赛设施

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。

竞赛现场符合消防安全规定，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。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、大赛关键环节与时间安排

（一）关键环节

1. 参赛选手报名

(1) 参赛队资格：中职组、高职组。参赛选手需为全日制在籍学生，性别不限。

(2) 组队要求：每个参赛队由 3名在籍同校学生组成。每校参赛队不多于 2 支。每个参赛队可配备指导教师 1-2 名。

(3) 人员变更：参赛选手和指导教师报名获得确认后不得随意更换。如备赛过程中参赛选手和指导教师因故无法参赛，须由所在院校于相应赛项开赛 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；团体赛选手因特殊原因不能参加比赛时，则视为自动放弃竞赛。

(4) 各参赛报名单位自行负责本单位参赛学生的资格审查工作，保存相关证明材料的复印件，以备查阅。

2. 熟悉场地

(8) 组委会安排在竞赛前1日下午后各参赛队统一有序的熟悉场地。

(9) 熟悉场地时严禁与现场工作人员进行交流，不发表没有根据以及有损大赛整体形象的言论。

(10) 熟悉场地严格遵守大赛各种制度，严禁拥挤，喧哗，以免发生意外事故。

3. 比赛入场

(1) 参赛选手按规定时间到达指定地点，必须携带参赛证件，进行检录、一次加密、二次加密等流程，最终确定工位，选手迟到 10 分钟取消比赛资格。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。参赛选手所需的硬件、软件和辅助工具统一提供，参赛队不得使用自带的任何有存储功能的设备，如手机、U 盘、移动硬盘等。参赛队在赛前领取比赛任务并进比赛工位，比赛正式开始后方可进行相关操作。

(2) 除严格规定的量具或其他物品外，参赛选手不允许携带任何通讯及存储设备、纸质材料等物品进入赛场，赛场内提供比赛必备用品。

4. 比赛过程

(1) 在比赛过程中，参赛选手如有疑问，应举手示意，现场裁判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，现场裁判、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，经裁判长确认，予以启用备用设备。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。

(2) 凡在竞赛期间内提前离开的选手，不得返回赛场。

(3) 选手进入赛场必须听从现场裁判人员的统一布置和指挥，首先需对比赛设备、选配部件、工量具等物品进行检查和测试，如有问题及时向裁判人员报告。

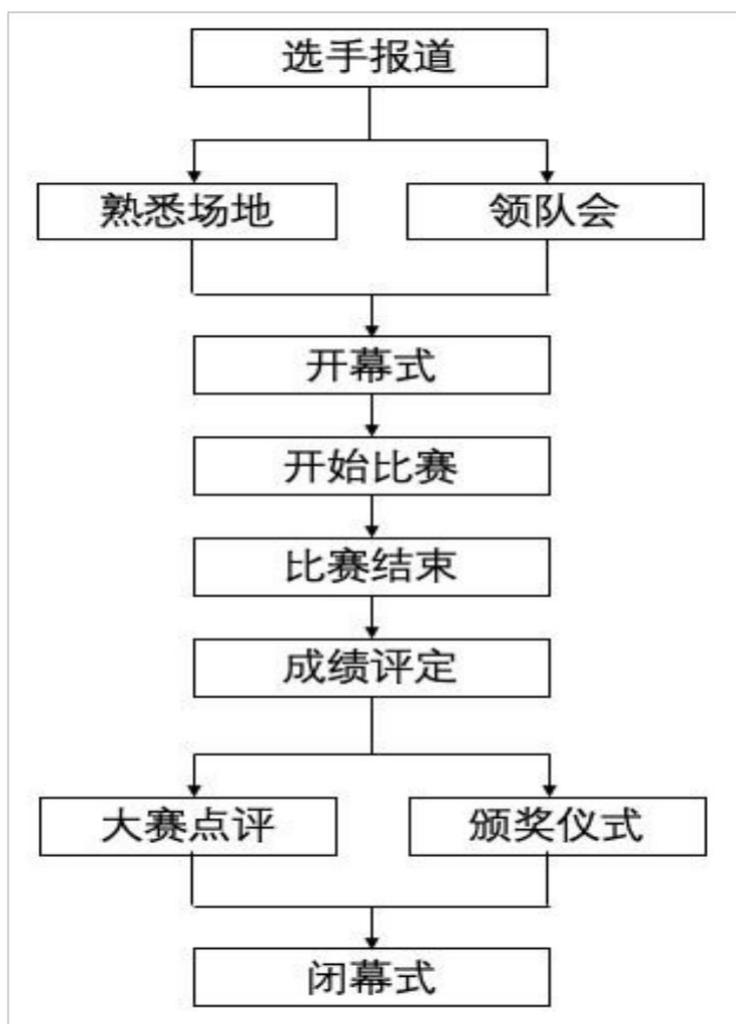
(4) 比赛过程中选手不得随意离开工位范围，不得与其它队伍的选手交流或擅自离开赛场。如遇问题时须举手向裁判员示意询问后处理，否则按作弊行为处理。

(5) 在比赛过程中只允许裁判员、工作人员进入现场，其余人员（包括指导教师）未经组委会同意不得进入赛场。

5. 比赛结束

比赛时间结束，选手应全体起立，结束操作。参赛选手要确认已成功提交竞赛要求的文档，裁判员与参赛选手一起签字确认，经工作人员查收清点所有文档后方可离开赛场，离开赛场时不得带走任何资料。

(二) 竞赛流程



(三) 时间安排

表 竞赛时间表

日期	时间	事项
竞赛前一日	12:00点前	各参赛队报到
	10:00—11:00	工作人员(含监考)培训会
	15:30—16:00	赛前领队会
	16:00—16:30	参赛队熟悉比赛场地

	17:00—18:00	现场裁判赛前检查，封闭赛场
竞赛当日	06:00—07:00	参赛队集合前往比赛现场
	07:00—07:10	赛场检录
	07:10—07:30	一次加密：参赛队抽取参赛编号
	07:30—07:45	二次加密：参赛队抽取赛位号
	07:45—08:00	参赛队进入比赛赛位，检查软硬件、题目发放
	08:00—12:00	竞赛进行
	12:00—12:20	收取各参赛队赛题及比赛结果文档
	12:00—14:00	申诉受理
	12:30—16:00	成绩评定与复核
	16:00—17:00	成绩汇总，报送及公布
	17:00—18:00	闭幕式

九、大赛评分标准制定原则、评分方法、评分细则及技术规范

（一）评分标准制定原则

1 竞赛评分制定严格遵守公平、公正的原则，大数据应用与服务赛项评分采用赛项结果评分方法，始终贯彻落实竞赛一贯坚持的公平、公正和公开原则。

2 参与竞赛成绩管理的组织机构包括裁判组、监督仲裁组等，裁判组实行“裁判长负责制”，监督仲裁组组成结构、人数和相关要求参建《

2023全国职业院校技能大赛监督仲裁工作管理办法》中监督仲裁遴选条件和办法。

3 裁判评分方法，根据评分标准，各项目评分裁判根据选手操作过程和操作结果进行评分，独立评分。

4 成绩产生方法。为保证公开、公平、公正、透明地进行成绩评定，在裁判员的评分中，取两个评分裁判平均分作为选手技能得分。

5. 成绩审核方法为各裁判员首先审核自身对选手的原始打分成绩，并签名，裁判长对所有裁判员的打分成绩进行审核，并签名，再由监督组对竞赛成绩抽检复核。

（二）评分方法

1 各竞赛模块分别计算得分，错误不传递，按规定比例计入团队总分。

2 竞赛满分为 100 分，根据各竞赛模块考核要求，制定评分细则进行评分，最终成绩按 100 分制进行排名。

3 在竞赛过程中，参赛选手如有不服从裁判裁决、扰乱赛场秩序和舞弊等行为的，由裁判长按照规定扣减相应分数，情节严重的将取消比赛资格，比赛成绩计 0 分。

4 参赛选手不得在比赛结果上标注含有本参赛队信息的记号，如有发现，取消奖项评比资格。

（三）评分细则(评分指标)

竞赛满分为100分，包括环境搭建（容器环境）、数据迁移和转换、数据挖掘、数据可视化、职业素养，具体评分细则如下：

模块		主要内容	任务时长	分值
模块一： 平台搭建 与运维	任务一： 大数据平台搭建	国产操作系统（麒麟Linux、阿里龙蜥操作系统）应用与运维，Hadoop 平台的安装部署和常用组件的安装部署。	120分钟	10
	任务二： 数据库配置维护	使用国产化数据库建库建表，运用基本的SQL语言完成数据的增删改查等操作。	120分钟	20
模块二： 数据获取 与处理	任务一： 数据获取与清洗	对CSV数据文件进行加载、清洗和转换等操作，识别和处理无效值，检查数据的一致性，将清洗后的数据保存到指定位置。	80分钟	10
	任务二： 数据标注	使用Python语言对数据进行分类标注。	80分钟	10
	任务三： 数据统计	基于Hadoop平台进行编译、打包、部署和执行程序，完成数据的统计工作。	80分钟	15
模块三： 业务分析 与可视化	任务一： 数据可视化	使用Web前端框架或者python可视化库对数据进行可视化展示。	120分钟	20
	任务二： 业务分析	报表分析，对大数据项目的业务场景和数据进行分析，撰写报告。	120分钟	10
职业素养		团队分工明确合理、操作规范、文明竞赛		5

（四）技术规范

本赛项的技术规范将包括：相关专业的教育教学要求、行业、职业技术标准，以及根据《职业教育专业目录（2021年）》、《职业教育专业简介》修订后的大数据相关专业人才培养标准和规范，适时地修订本赛项遵循的技术规范。

1 基础标准

标准号/规范简称	名称
GB/T11457-2006	信息技术、软件工程术语

GB8566-88	计算机软件开发规范
GB/T12991-2008	信息技术数据库语言SQL第1部分：框架
GB/T21025-2007	XML使用指南
GB/T28821-2012	关系数据管理系统技术要求
LD/T81.1-2006	职业技能实训和鉴定设备技术规范

2 大数据技术相关标准

标准号/规范简称	名称
GB/T38672-2020	信息技术大数据接口基本要求
GB/T38673-2020	信息技术大数据系统基本要求
GB/T38676-2020	信息技术大数据存储与处理系统功能测试要求
GB/T38643-2020	信息技术大数据分析系统功能测试要求
GB/T38675-2020	信息技术大数据计算系统通用要求
GB/T38633-2020	信息技术大数据系统运维和管理功能要求
职业编码：2-02-10-11	国家职业技术技能标准大数据工程技术人员
标准代码：510062	大数据工程化处理与应用职业技能等级标准
标准代码：510010	大数据平台运维职业技能等级标准

3 软件开发标准

标准号/规范简称	名称
GB/T8566-2001	信息技术软件生存周期过程
GB/T8566-2001	软件支持环境
GB/T14079-1993	软件维护指南
GB/T17544-1998	信息技术软件包质量要求和测试
GB/T8567-2006	计算机软件文档编制规范
SJ/T11291-2003	面向对象的软件系统建模规范

十、大赛平台说明

（一）竞赛设备

符合技术要求的各类设备的技术参数、工装器具的技术规格、软件版本号等信息如下：

序号	设备名称	备注
1	服务器集群	支撑信创大数据技术竞赛平台运行，可以采用云端方式，每台 配置如下： CPU：不少于32核 内存：不少于128GB 存储：不少于480GB SSD系统盘，总容量不少于4TB存储盘
2	信创大数据竞赛平台（海彦科技）	基于PaaS（容器云）、云原生开发等先进技术构建实训云平台，是面向云原生大数据平台，可训练贯穿大数据生命周期核心知识和技能，包括环境搭建和管理、数据迁移和转换、数据存储及管理、数据分析及挖掘、数据展现和应用等。
3	PC机	各参赛选手比赛使用电脑，采用PC机，每台配置如下： CPU：8核或性能不低于i5处理器 内存：不少于8GB 硬盘：不少于200GB 显示器：分辨率1024*768以上 操作系统：麒麟Linux
4	交换机	为PC机提供网络支撑。 1U机架式交换机。 网口：不少于24个自适应千兆电口。

（二）软件环境

赛项需要软件环境如下表所示：

设备类型	软件类别	软件名称、版本号
服务器	大数据集群操作系统	麒麟Linux、阿里龙蜥操作系统
	大数据分析平台组件	Hadoop 3.3.2
		Yarn 3.3.2
		Zookeeper 3.5.7
		Hive 3.1.2
		Flume 1.9.0
		HBase 2.4.11
		Sqoop 1.4.6
		kafka 2.12-2.8.0
		Spark 3.1.3
		Flink 1.13.0
	数据库	国产数据库 MySQL
PC机	PC 操作系统	麒麟Linux
	浏览器	火狐
	开发语言	Java 8
		Python 3
	项目管理工具	Maven 3.8.4
	开发工具	HBuilderX 3.7.10
		IDEA 2023 (Community Edition)
		PyCharm 2023 (Community Edition)
文档编辑器	WPS 2019 及以上	
输入法	拼音输入法	

十一、大赛安全保障

(一) 组织机构

在赛项执行委员会领导下，成立赛事保障组，负责赛项安全，成员包括承办院校主抓安全的校领导、学生工作处、后勤处、保卫处、合作企业技术工程师等相关人员。

(二) 比赛环境

- 1 赛场的布置，赛场内的器材、设备，应符合国家有关安全规定。
- 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。
- 3 承办院校应提供保障应急预案实施的条件，明确制度和预案。
- 4 赛项执委会须会同承办院校制定开放赛场和体验区的人员疏导方案。
- 5 大赛期间，赛项承办院校须在赛场设置医疗救护工作站。
6. 参赛选手、赛项裁判、工作人员严禁携带通讯、摄录设备和未经许可的记录用具进入比赛区域。

（三）生活条件

- 1 比赛期间，原则上由赛项承办院校统一安排参赛选手和指导教师食宿。承办院校须尊重少数民族参赛人员的宗教信仰及文化习俗。
- 2 比赛期间安排的住宿场所应具有旅游业经营许可资质。
- 3 大赛期间组织的参观和观摩活动的交通安全由赛区组委会负责。

（四）参赛队职责

- 1 各参赛单位在组织参赛队时，须安排为参赛队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。
- 3 各参赛队伍须加强对参与比赛人员的安全管理，实现与赛场安全管理的对接。
- 4 参赛队车辆，一律凭赛项执委会核发的证件出入校门，并按指定线路行驶，按指定地点停放。

（五）应急处理

比赛期间发生意外事故，发现者应第一时间报告赛项执行委员会，同时采取措施避免事态扩大。赛项执行委员会应立即启动预案予以解决并报告大赛组委会。赛项出现重大安全问题可以停赛，是否停赛由大赛组委会决定。事后，赛项执行委员会应向大赛组委会报告详细情况。

十二、大赛组织与管理

（一）大赛组织

1 成立执行委员会：在大赛组委会的领导下，成立“信创大数据技术应用”赛项执行委员会，全面负责赛项的具体组织与实施。

2 技术专家组：负责技术层面的决策和指导，确保比赛的技术要求和标准得到准确执行。

3 裁判评判组：由行业专家组成，确保比赛的公正性和公平性，同时负责成绩的汇总和发布。

4 监督仲裁组：负责监督比赛的各个环节，处理参赛队伍的申诉，确保比赛的顺利进行。

（二）大赛设备与设施管理

1 设备采购与配置

(1) 根据大赛的技术要求和参赛人数，提前采购足够数量的信创大数据相关设备，包括但不限于服务器、存储设备、网络设备、终端计算机等。

(2) 确保所有设备符合信创标准，并能够满足参赛者在比赛中进行大数据处理、分析和应用的需求。

(3) 为每台设备配置相应的操作系统、大数据处理工具和其他必要软件。

2 设备安装与调试

(1) 在比赛场地内设立专门的设备安装区域，按照比赛要求对所有设备进行安装和调试。

(2) 在安装过程中，确保所有设备之间的网络连接稳定，数据传输速度达到比赛要求。

(3) 对所有设备进行全面的测试，确保其在比赛期间能够稳定运行。

3 设施保障

(1) 提供稳定的电力供应，确保比赛期间不断电。

(2) 安装空调和通风设备，确保比赛场地内的温度和湿度适宜，为参赛者提供舒适的比赛环境。

(3) 设立专门的休息区，为参赛者提供必要的饮食和休息设施。

4 设备维护与技术支持

(1) 比赛期间，设立专门的设备维护和技术支持团队，随时解决参赛者在比赛过程中遇到的设备问题。

(2) 提供24小时的技术支持服务，确保参赛者在需要时能够得到及时的帮助。

十三、附件

背景描述

在当今数字化飞速发展的时代，大数据技术在各行各业都扮演着至关重要的角色，而在学校数据管理领域，大数据的应用更是愈发凸显其重要性。学校作为一个庞大而复杂的组织体系，拥有众多的学生、教职工和各种资源，数据的产生和管理变得日益庞大而复杂。因此，大数据技术的引入成为学校数据治理的一项重要举措。

在职业教育领域，职业大数据比赛的举办对学生、学校和整个教育体系都具有重要的意义。首先，大数据比赛为职业教育学生提供了一个实践和展示自己技能的平台。通过参与比赛，学生可以运用所学的大数据技术知识，解决实际问题，提高实际操作能力。这不仅有助于加深学生对大数据领域的理解，还锻炼了他们的团队协作和问题解决能力。

其次，职业大数据比赛有助于激发学生对科学技术的兴趣。通过实际操作和创新性的比赛内容，学生更容易对大数据技术产生浓厚的兴趣，并在实践中感受到这一领域的魅力。这种积极的学习体验有助于激发学生对科技领域的兴趣，培养创新意识，为其未来的职业发展奠定基础。

另外，大数据比赛也为学校提供了一个展示自身教学水平和培养优秀学生的机会。优异的比赛成绩不仅可以提升学校的声誉，还有助于吸引更多学生报考相关专业。同时，学校通过组织和支持学生参与大数据比赛，也能够加深与企业、行业的合作关系，促进校企合作的深度发展。

总体而言，职业大数据比赛在推动学生综合素质提升、促进科技创新、推动学校发展等方面发挥着重要的作用。通过这样的比赛活动，职业教育能够更好地适应社会需求，培养更符合产业发展需求的人才，为学生未来的职业生涯打下坚实基础。

模块一：平台搭建与运维

任务一：大数据平台搭建

子任务一：Hadoop 完全分布式安装配置

本任务需要使用 root 用户完成相关配置。打开竞赛平台的赛题资源的容器资源，利用智联助手的SSH集成窗口打开容器，基于竞赛平台进行完全分布式模式Hadoop的搭建和管理。相关安装文件在容器“/opt”目录下，请选择对应的安装包进行安装，用不到的可忽略。主机名为master、slavel、slave2的三台节点都需要安装JDK和Hadoop。

(1) 在master节点执行命令，创建jdk安装目录“/data/jdk”和Hadoop安装目录“/data/hadoop”，执行ls命令查看创建的目录。提交创建目录的命令和查看的结果截图；

(2) 在master节点执行tar命令（请确保在解压缩文件时，不要保留原始文件路径中的第一个目录层级），将节点中“/opt”目录下jdk安装文件解压到节点中的“/data/jdk”目录，将Hadoop安装文件解压到节点中的“/data/hadoop”目录，执行ls命令分别查看解压后的文件。提交解压命令和查看的结果截图；

(3) 在三台节点分别修改“/etc/hosts”配置，添加三台节点ip和主机名映射；配置 ssh 免密，实现master、slavel、slave2三台节点间相互免密登录，在master节点执行“ssh slavel”命令和在slave2节点执行“ssh master”命令验证。提交“/etc/hosts”配置截图、在master执行的免密命令和执行结果截图、验证命令的结果截图；

(4) 在master节点使用scp命令并使用绝对路径将jdk目录拷贝到slavel和slave2（若路径不存在，则需新建），并在三台节点的“/etc/profile”文件中配置jdk环境变量并使其生效，配置完毕后，在节点中执行“java -version”命令，检测jdk是否安装成功。提交scp命令、环境变量截图和使环境变量生效命令、检测命令结果截图；

(5) 依次配置hadoop-env.sh、core-site.xml、workers配置文件，其中NameNode的地址为master节点，端口为9000，master、slave1、slave2节点均作为DataNode，配置好相关环境。在master节点使用scp命令将配置完的Hadoop安装目录直接拷贝至slave1和slave2（若路径不存在，则需新建），在三台节点的“/etc/profile”文件中配置Hadoop环境变量，配置完毕后，在容器中执行“hadoop version”命令，检测Hadoop是否安装成功。提交修改的配置内容截图、scp命令、环境变量截图、检测命令结果截图；

(6) 在master节点初始化Hadoop环境namenode。提交初始化命令和结果截图；

(7) 在master节点依次启动HDFS、YARN集群，在三个节点分别执行jps命令，查看节点中的进程。提交jps查看结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务二：Hive 安装配置

本任务需要使用 root 用户完成相关配置，已安装Hadoop及需要配置前置环境。具体要求如下：

(1) 在master节点中执行命令，创建Hive安装目录“/data/hive”，执行ls命令查看创建的目录。提交创建目录的命令和查看的结果截图；

(2) 在容器中执行tar命令（请确保在解压缩文件时，不要保留原始文件路径中的第一个目录层级），将容器中“/opt”目录下Hive安装文件解压到容器中的“/data/hive”目录，在“/etc/profile”文件中配置Hive环境变量并使其生效。提交解压命令和环境变量内容截图；

(3) 使用schematool命令初始化源数据库的元数据。提交命令和结果截图；

(4) 在启动hadoop集群的情况下，使用CLI启动Hive，进入Hive客户端；在Hive默认数据库下创建一个名为student的管理表，字段有id（int），name（string），并通过insert语句往student表中插入一条测试数据。提交创建表命令和执行结果截图、insert语句和执行成功的结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

任务二：数据库配置维护

子任务一：数据库配置

打开竞赛平台的赛题资源的数据库资源，利用智联助手的SSH集成窗口打开master节点，基于竞赛平台进行MySQL的搭建和管理。具体要求如下：

(1) 在赛题资源的数据库资源对应的ip的节点中执行命令，创建MySQL安装目录“/data/mysql”；执行tar命令，将容器中“/opt”目录下Mysql安装文件解压到“/data/mysql”目录，执行ls命令查看解压后的文件。提交tar命令和ls结果截图；

(2) 安装好MySQL后，对数据库进行初始化，将\$MYSQL/bin目录下的mysql脚本创建软链到/usr/bin目录，在/etc/systemd/system目录中配置mysql.service，配置开机自启动并启动数据库。提交初始化命令、创建软链命令、mysql.service配置内容截图、配置自启动命令、启动数据库结果截图；

(3) 使用root用户和密码登录MySQL，然后将root用户的密码修改为赛题资源的数据库资源的密码，配置授予root用户在任何主机（'%'）

上连接时对所有数据库 (*.*) 的全部权限，并允许该用户授予权限给其他用户，配置完刷新权限。提交修改密码sql语句和结果截图、配置授予权限sql语句和结果截图；

(4) 连接数据库，先用sql创建一个“exam”数据库(字符集设置utf8mb4)，再创建一个“test”新用户，密码与root密码一致，将“exam”数据库权限授予“test”用户，刷新权限后用“test”账号登录数据库，执行“show databases;”命令查看数据库。提交创库sql语句、建用户语句、授权语句、“show databases;”的结果截图；

(5) 将“test”用户删除。提交删除语句；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务二：数据库表维护

打开竞赛平台的赛题资源的数据库资源，利用智联助手的数据库工具打开“exam”数据库，基于竞赛平台进行MySQL的数据库表维护。

(1) 在MySQL数据库的“exam”数据库中创建“athlete_events”数据表。数据表的表字段格式如下：

表1 “athlete_events” 的表字段结构

字段	类型	注释
Name	varchar	名称
Sex	varchar	性别
Age	int	年龄
Height	int	身高
Weight	int	体重
Games	varchar	赛事

City	varchar	城市
Event	varchar	比赛项目
Medal	varchar	奖牌

提交建表sql语句和结果截图；

(2) 为“athlete_events”表添加数据记录。记录如下：

Name	Sex	Age	Height	Weight	Games	City	Event	Medal
Cao Yuan	M	21	160	42	2016 Summer	Rio de Janeiro	Diving Men's Springboard	Gold
Chen Aisen	M	20	168	60	2016 Summer	Rio de Janeiro	Diving Men's Platform	Gold
Chen Long	M	27	188	81	2016 Summer	Rio de Janeiro	Badminton Men's Singles	Gold
Chen Peina	F	27	172	63	2016 Summer	Rio de Janeiro	Sailing Women's Windsurfer	Silver
Chen Dequan	M	18	176	66	2014 Winter	Sochi	Short Track Speed Skating Men's 5,000 metres Relay	Bronze
Jia Zongyong	M	22	175	70	2014 Winter	Sochi	Freestyle Skiing Men's Aerials	Bronze
Cai Yun	M	32	181	68	2012 Summer	London	Badminton Men's Doubles	Gold
Zou Kai	M	24	158	55	2012 Summer	London	Gymnastics Men's Team All-Around	Gold
Zhou Lulu	F	24	175	133	2012 Summer	London	Weightlifting Women's Super-Heavyweight	Gold
Zhang Jike	M	24	178	70	2012 Summer	London	Table Tennis Men's Singles	Gold
Guo Xinxin	F	26	157	60	2010 Winter	Vancouver	Freestyle Skiing Women's Aerials	Bronze
Li Nina	F	27	160	52	2010 Winter	Vancouver	Freestyle Skiing Women's Aerials	Silver
Bao Yingying	F	24	172	67	2008 Summer	Beijing	Fencing Women's Sabre, Team	Silver
Feng Kun	F	29	183	75	2008 Summer	Beijing	Volleyball Women's Volleyball	Bronze

Han Xiaoping	M	22	173	75	2006 Winter	Torino	Freestyle Skiing Men's Aerials	Gold
Chen Jing	F	28	182	75	2004 Summer	Athina	Volleyball Women's Volleyball	Gold
Chen Zhong	F	21	183	73	2004 Summer	Athina	Taekwondo Women's Heavyweight	Gold
Zhao Hongbo	M	28	177	75	2002 Winter	Salt Lake City	Figure Skating Mixed Pairs	Bronze
Cai Yalin	M	23	174	60	2000 Summer	Sydney	Shooting Men's Air Rifle, 10 metres	Gold
Chen Xiaomin	F	23	158	63	2000 Summer	Sydney	Weightlifting Women's Middleweight	Gold
Chen Lu	F	21	162	52	1998 Winter	Nagano	Figure Skating Women's Singles	Bronze
He Qi	F	22	178	68	1996 Summer	Atlanta	Volleyball Women's Volleyball	Silver
Lu Lin	M	27	174	60	1996 Summer	Atlanta	Table Tennis Men's Doubles	Silver
Li Xiaoshuang	M	22	157	52	1996 Summer	Atlanta	Gymnastics Men's Individual All-Around	Gold
Ye Qiaobo	F	29	168	70	1994 Winter	Lillehammer	Speed Skating Women's 1,000 metres	Bronze

提交第一条数据的插入sql语句及所有数据添加结果截图；

(3) 新增“Year”字段(int)，注释为“年份”，将“Games”字段中的年份数字提取到“Year”字段，并查询“Year”在2000年到2008年之间的数据。提交新增字段sql和结果截图、提取年份sql、查询的sql和结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务三：数据表查询

(1) 筛选出比赛项目为“Volleyball Women’s Volleyball”且身高大于“180”的数据，并按照年份“Year”进行升序排序。提交sql和结果截图；

(2) 查询2016年参赛人员年龄、身高和体重各自的最大值、最小值和平均值。提交sql和结果截图；

(3) 按年份分组查询每年参赛人员的男女人数各多少。提交sql和结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

模块二：数据获取与处理

任务一：数据获取与清洗

子任务一：数据获取

打开竞赛平台赛题资源的文件资源的P1-M2-T1文件分类，下载parse_athlete_events.py文件。赛题资源中静态资源athlete_events.html是奥运会运动员数据和奖牌成绩列表内容，点击进入可打开网页。parse_athlete_events.py为Python脚本文件，程序读取athlete_events.html，使用lxml对网页进行解析，提取相应的列表数据，并将结果输出。

(1) 用竞赛平台智联助手的python脚本工具打开parse_athlete_events.py文件，补全文件中【1】代码，配置公共资源地址url。提交url地址代码；

(2) 使用浏览器打开赛题资源中静态资源文件athlete_events.html网页文件，通过“审查”工具进行网页结构分析。补全parse_athlete_events.py中【2】代码，实现获取奥运会运动员数据和奖牌成绩div列表。提交补全代码；

(3) 补全parse_athlete_events.py中【3】~【5】代码，实现“text”，“举办城市”和“代表队伍”文本内容提取。提交补全代码；

(4) 运行parse_athlete_events.py脚本，完成奥运会运动员数据和奖牌成绩列表的解析并打印输出。提交运行结果截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务二：数据处理

打开竞赛平台赛题资源的文件资源的P1-M2-T1文件分类，下载athlete_events.csv文件。通过编写python代码完成对相关数据文件中数据的清洗和整理。请分析相关数据集，根据题目规定要求实现数据处理，具体要求如下：

(1) 查看athlete_events.csv中数据总数、标准差、均值、最小值、四分之一分位数、二分之一分位数、四分之三分位数和最大值。提交代码截图和结果截图；

(2) 对athlete_events.csv文件进行处理，对于Age列数据，采用“Age”填充缺失值，并将年龄数字大于等于90的数据删除，然后存入athlete_events_c1.csv中。提交代码截图；

(3) 对athlete_events_c1.csv文件进行处理，对于Height列和Weight列数据，将缺失值的数据删除，然后存入athlete_events_c2.csv中。提交代码截图；

(4) 对athlete_events_c2.csv文件进行处理，对于Weight列数据，小数点后超过一位数的数据，将小数四舍五入改为一位数，然后存入athlete_events_c3.csv中。提交代码截图；

将上述任务的命令和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

任务二：数据标注

对athlete_events_c3.csv进行标注，判断运动员是否获得奖牌，具体的标注规则如下：

(1) 如果“Medal”列数据不为空，则数据标注为“yes”；

(2) 如果“Medal”列数据为空，则数据标注为“no”；

标注好的数据存储为列“Have_Medal”并和athlete_events_c3.csv数据合并存入result.csv。提交代码截图和csv文件；

将上述任务的结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

任务三：数据统计

子任务一：处理异常数据

打开竞赛平台赛题资源的文件资源的P1-M2-T3文件分类，下载TravelInsurancePrediction.csv文件。该文件存储了旅游保险预测相关数据，数据中有以下内容：

Age: 年龄，int类型

EmploymentType: 职业种类，varchar类型

GraduateOrNot: 是否毕业生，varchar类型

AnnualIncome: 年收入，int类型

FamilyMembers: 家庭成员，int类型

ChronicDiseases: 是否有慢性病，int类型

FrequentFlyer: 是否经常坐飞机，varchar类型

EverTravelledAbroad: 曾经出国旅游，varchar类型

TravelInsurance: 是否购买旅游保险，int类型

编写MapReduce程序，实现以下功能：清除曾经有出国旅游的人员数据，在控制台按顺序打印输出前20条数据，输出结果到HDFS，使用命令查看结果。提交代码截图和结果截图；

将上述任务的程序和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务二：数据统计

使用竞赛平台赛题资源的文件资源的P1-M2-T3文件分类的TravelInsurancePrediction.csv文件，编写MapReduce程序，实现以下功能：根据“GraduateOrNot”字段统计毕业生与非毕业生人数，并在控制台输出，输出结果到HDFS，使用命令查看结果。提交代码截图和结果截图；

将上述任务的程序和结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

模块三：数据分析与可视化

任务一：数据分析与可视化

使用竞赛平台可视化工具或者下载可视化操作文件至本地进行代码编写并以图表展示。

子任务一：柱状图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具，点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源P1-M3-T1-SUBT1文件分类，下载visualization.zip文件至本地解压，编写补充代码，实现Web网页形式对2008年得奖数量前十的国家数据进行可视化展示：

【竞赛平台的数据分析与可视化工具】

(1) 编写补充yAxis对象，获取相关数据中barData数据，设置y轴显示类型为“类目轴”、设置坐标轴文字颜色值为：#999999，大小为：

12、设置坐标轴在 `grid` 区域中的分隔线颜色为：`#CAD3E0`，线的类型为：虚线、设置y轴显示数据为“2008年得奖数量前十的国家名称”。

(2) 编写补充`series`对象，获取相关数据中`barData`数据，设置图表显示类型为柱状图、设置系列名称为“奖牌数量”、设置柱条的宽度为20，背景颜色为：`rgba(180, 180, 180, 0.2)`、将“2008年得奖数量前十的国家”，`barData`对象中的数据设置为柱状图显示数据。

(3) 附上“2008年得奖数量前十的国家柱状图”截图与相关代码截图。

【本地操作】

解压 `visualization.zip` 文件至本地，根据 `visualization/data/data.js` 文件中`barData`对象中的数据，补充完整 `visualization/js/chart.js` 文件中`getBarChart()` 函数的代码

(1) 编写补充`yAxis`对象，获取`barData`数据，设置y轴显示类型为“类目轴”、设置坐标轴文字颜色值为：`#999999`，大小为：12、设置坐标轴在 `grid` 区域中的分隔线颜色为：`#CAD3E0`，线的类型为：虚线、设置y轴显示数据为“2008年得奖数量前十的国家名称”。

(2) 编写补充`series`对象，获取`barData`数据，设置图表显示类型为柱状图、设置系列名称为“奖牌数量”、设置柱条的宽度为20，背景颜色为：`rgba(180, 180, 180, 0.2)`、将“2008年得奖数量前十的国家”，`barData`对象中的数据设置为柱状图显示数据。

(3) 运行网页，附上“2008年得奖数量前十的国家柱状图”截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务二：折线图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具，点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源P1-M3-T1-SUBT2文件分类，下载visualization.zip文件至本地解压，编写补充代码，实现Web网页形式对中国在各届夏季奥运会上的奖牌数量变化数据进行可视化展示。任务点如下：

【竞赛平台的数据分析与可视化工具】

(1) 编写补充tooltip对象，设置提示框组件的触发类型为坐标轴触发、设置指示器类型为：直线指示器、设置提示框浮层的文字颜色：#666666，字体大小为：12。

(2) 编写补充xAxis对象，获取相关数据中线Data数据，设置X轴显示类型为“类目轴”、设置坐标文字显示为:#999999，文字大小设置为：12，文字间隔为0，文字倾斜角度为40度、设置X轴显示坐标为“各界夏季奥运会”。

(3) 编写补充series对象，获取相关数据中线Data数据，设置图表显示类型为'line'、设置线条显示平滑，标记大小为6的三角形、设置折线图文字显示(将文字颜色设置为#999999，文字大小设置为：12)、将“中国在各届夏季奥运会上的奖牌数量”对象中的数据设置为折线显示数据。

(4) 附上“中国在各届夏季奥运会上的奖牌数量变化折线图”截图与相关代码截图。

【本地操作】

解压 visualization.zip 文件至本地，根据 visualization/data/data.js 文件中lineData对象中的数据，补充完整 visualization/js/chart.js 文件中getLineChart()函数的代码

(1) 编写补充tooltip对象，设置提示框组件的触发类型为坐标轴触发、设置指示器类型为：直线指示器、设置提示框浮层的文字颜色：#666666，字体大小为：12。

(2) 编写补充xAxis对象，获取lineData数据，设置X轴显示类型为“类目轴”、设置坐标文字显示为:#999999，文字大小设置为：12，文字间隔为0，文字倾斜角度为40度、设置X轴显示坐标为“各界夏季奥运会”。

(3) 编写补充series对象，获取lineData数据，设置图表显示类型为'line'、设置线条显示平滑，标记大小为6的三角形、设置折线图文字显示(将文字颜色设置为#999999，文字大小设置为：12)、将“中国在各届夏季奥运会上的奖牌数量”对象中的数据设置为折线显示数据。

(4) 运行网页，附上“中国在各届夏季奥运会上的奖牌数量变化折线图”截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务三：饼图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具，点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源P1-M3-T1-SUBT3文件分类，下载visualization.zip文件至本地解压，编写补充代码，实现Web网页形式对2016夏季奥运会运动员年龄分布数据进行可视化展示。任务点如下：

【竞赛平台的数据分析与可视化工具】

(1) 编写补充legend对象，获取相关数据中pieData数据，设置图例的朝向为：垂直显示、设置图例在X轴方向上的位置为右、设置图例上

显示的文字信息为：年龄段，图例项的图标都设置为圆形、设置图例文字颜色为：#999999，大小为：12。

(2) 编写补充series对象，获取相关数据中pieData数据，设置图表类型为'pie'、设置系列名称为'2016夏季奥运会运动员年龄分布'、设置饼图半径为['30%', '70%']、设置饼图高亮状态，标签文字颜色：#999999，大小：24，居中显示、将'2016夏季奥运会运动员年龄分布'对象中的数据设置为饼图显示数据。

(3) 附上“2016夏季奥运会运动员年龄分布饼图”截图与相关代码截图。

【本地操作】

解压 visualization.zip 文件至本地，根据 visualization/data/data.js 文件中 pieData 对象中的数据，补充完整 visualization/js/chart.js 文件中 getPieChart() 函数的代码

(1) 编写补充legend对象，获取pieData数据，设置图例的朝向为：垂直显示、设置图例在X轴方向上的位置为右、设置图例上显示的文字信息为：年龄段，图例项的图标都设置为圆形、设置图例文字颜色为：#999999，大小为：12。

(2) 编写补充series对象，获取pieData数据，设置图表类型为'pie'、设置系列名称为'2016夏季奥运会运动员年龄分布'、设置饼图半径为['30%', '70%']、设置饼图高亮状态，标签文字颜色：#999999，大小：24，居中显示、将'2016夏季奥运会运动员年龄分布'对象中的数据设置为饼图显示数据。

(3) 运行网页，附上“2016夏季奥运会运动员年龄分布饼图”截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务四：散点图数据分析与可视化

使用竞赛平台数据治理下数据分析与可视化工具，点击柱状图编写补充代码或者打开竞赛平台赛题资源的文件资源P1-M3-T1-SUBT4文件分类，下载visualization.zip文件至本地解压，编写补充代码，实现Web网页形式对美国各届夏季奥运会奖牌数量数据进行可视化展示。任务点如下：

【竞赛平台的数据分析与可视化工具】

(1) 编写补充tooltip对象，设置提示框组件的触发类型为数据项图形触发、设置指示器类型为：十字准星指示器、设置提示框浮层的文字颜色：#666666，字体大小为：12。

(2) 编写补充xAxis对象，获取相关数据中scatterData数据，设置X轴，显示类型为“类目轴”、设置坐标文字显示为:#999999，文字大小设置为：12、设置X轴显示坐标为“各届夏季奥运会”。

(3) 编写补充series对象，获取相关数据中scatterData数据，设置图表显示类型为'scatter'、设置标记大小为6的三角形、设置散点图图文字显示(显示标签, 标签的位置设为图形顶部，文字颜色设置为#999999、文字大小设置为：12)、将“美国各届夏季奥运会奖牌数量”对象中的数据设置为散点图显示数据。

(4) 附上“美国各届夏季奥运会奖牌数量散点图”截图与相关代码截图。

【本地操作】

解压 visualization.zip 文件至本地，根据 visualization/data/data.js 文件中 scatterData 对象中的数据，补充完整 visualization/js/chat.js 文件中 getScatterChart() 函数的代码

(1) 编写补充 tooltip 对象，设置提示框组件的触发类型为数据项图形触发、设置指示器类型为：十字准星指示器、设置提示框浮层的文字颜色：#666666，字体大小为：12。

(2) 编写补充 xAxis 对象，获取 scatterData 数据，设置 X 轴，显示类型为“类目轴”、设置坐标文字显示为：#999999，文字大小设置为：12、设置 X 轴显示坐标为“各届夏季奥运会”。

(3) 编写补充 series 对象，获取 scatterData 数据，设置图表显示类型为‘scatter’、设置标记大小为 6 的三角形、设置散点图图文字显示(显示标签, 标签的位置设为图形顶部，文字颜色设置为 #999999、文字大小设置为：12)、将“美国各届夏季奥运会奖牌数量”对象中的数据设置为散点图显示数据。

(4) 运行网页，附上“美国各届夏季奥运会奖牌数量散点图”截图与相关代码截图。

将上述任务的代码和结果提交到竞赛平台的赛题结果上传中对应的任务序号下。

任务二：数据分析

使用电子表格软件进行查询统计并使用图表进行展示。

子任务一：Excel 数据柱状图分析

打开竞赛平台赛题资源的文件资源的 P1-M3-T2 文件分类，下载 ANALYSE.xlsx 文件。使用电子表格软件打开 ANALYSE.xlsx 文件，对表格内的数据进行操作分析，并最终使用电子表格软件自带的图表功能，展示 2016 年奥运会获奖牌数量前十的 Team 柱状图，任务点如下：

- (1) 统计每个“Team”的“奖牌数量”；
- (2) 对“Team”的“奖牌数量”进行降序排列；
- (3) 取“Team”的“奖牌数量”前十使用柱状图进行显示；
- (4) 根据第1名奖牌数占前10名总奖牌数的比例，以及第1、2、3名奖牌差距，写一段分析报告。

将上述任务结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

子任务二：Excel数据折线图分析

打开竞赛平台赛题资源的文件资源的P1-M3-T2文件分类，下载ANALYSE.xlsx文件。使用电子表格软件打开ANALYSE.xlsx文件，对表格内的数据进行操作分析，并最终使用电子表格软件自带的图表功能，展示2016年奥运会运动员年龄小于30岁人数数量折线图，任务点如下：

- (1) 统计每个“Age”的“数量”；
- (2) 筛选“Age”小于30的“数量”，并按照年龄升序使用折线图进行显示；
- (3) 根据“Age”与“数量”的变化，写一段分析报告。

将上述任务的结果截图提交到竞赛平台的赛题结果上传中对应的任务序号下。

模块四：职业素养

职业大数据比赛职业素养评分关注团队合作、创新、问题解决、沟通、项目管理、职业操守等方面。参赛者需展现学科知识运用、自主学习、文档撰写能力。全面素养将在比赛中为团队成功贡献关键因素。

任务要求：

- (1) 团队合作：能力在团队中合作协调，有效沟通，共同完成项目任务。评价团队成员之间的合作默契、协同工作的能力。

(2) 创新能力：考察参赛者在解决问题时的创造性思维和创新能
力，包括提出独特的解决方案、采用新方法和技术等。

(3) 问题解决能力：能够迅速识别问题，采用合适的方法解决问
题，具备分析和解决实际问题的能力。

(4) 沟通能力：能够清晰、准确地表达自己的观点，有效地与团
队成员和评审沟通，包括书面报告、口头演讲等。

(5) 文档撰写能力：能够撰写清晰、完整、规范的文档。