

中国国际数字和软件服务创新大赛
产业大数据应用设计与开发赛项
技术文件

一、竞赛目标

本赛项是在国家大力推进新一代人工智能发展规划和《第十个五年规划和 2035 年远景目标纲要》新时代背景下开展的一项针对大数据技术应用的赛事。赛项具体针对大数据工程师岗位，以工程应用为项目背景进行设计，以大数据应用技术为主要考查方向的赛事，重点考查参赛选手在大数据处理与分析方面的能力。这包括对数据分析计算、数据可视化等方面的能力，以及对数据处理技术工具和框架等方面的工程应用能力。

赛项由“大数据分布式分析计算”、“大数据可视化”二个任务组成。全面考察选手对大数据技术的设计和开发能力。

二、竞赛内容

竞赛内容基于给定的产业大数据项目课题，要求参赛队完成以下两个任务。

任务 1：大数据分布式分析计算

本任务的目的是考核参赛队在大数据分布式环境下的开发能力。根据赛题信息，完成业务分析与任务理解，使用给定的大数据基础环境、配置工具，完成大数据分布式环境下的开发工作。

(1) 数据的预处理

对整理好的数据进行预处理，例如查询原始数据维度特征，或者在原始数据不完整的情况下，有可能处理其中的缺失值、重复值、异常值和错误值等。

(2) 数据分析与计算

参赛队基于给定的样例数据集,将大规模数据划分为多个数据块,每个数据块存储在不同的节点上,以便于并行处理。分布式数据存储:通过 Hadoop 等分布式平台对数据进行存储和管理。分布式数据处理:通过 MapReduce 等分布式计算框架进行分布式处理的实现。

任务 2 : 大数据可视化

本任务的目的是考核参赛队的数据可视化能力。要求参赛队通过相关技术对给定的数据分布式计算结果进行 API 调用,然后对这些数据进行可视化展示。

(1) 搭建 Web 应用系统

参赛队需要搭建 Web 应用系统,应用系统能够获取大数据分析后的结果数据,并提供对外数据访问和呈现的能力。

(2) 实现大数据可视化

参赛队需要将数据可视化报告呈现出来,可采用柱状图、饼图、折线图等方式,直观地展示计算分析过程中的数据变化趋势和性能表现情况等,并对数据可视化结果进行详细解释和分析。

三、竞赛方式

(一) 竞赛形式

初赛采取线上比赛形式。决赛采取线下比赛形式。

(二) 组队方式

赛项为团队赛。每支参赛队由 3 名参赛选手组成,每队可配 1-2 名指导教师,指导教师须为本校专兼职教师,参赛选手和指导教师报名获得确认后不得随意更换。

参赛选手须为全日制普通高等院校在籍学生。（以报名时的学籍信息为准）。参赛选手年龄须不超过 26 周岁，年龄计算的截止时间以竞赛当年规定为准。

四、竞赛流程

（一）初赛竞赛流程

初赛通过东软在线考试系统进行。具体时间、平台地址和账号待报名结束后统一发放。根据参赛队数量，初赛可能采取分区、多场进行。

（二）决赛竞赛流程

决赛的竞赛流程在初赛完成后发出。

五、技术规范

（一）标准与规范

系统梳理了大数据领域的国际、国家、行业技术以及职业资格标准。

序号	标准号	标准名称
1	2-02-10-11	大数据工程技术人员
2	GB/T 35295-2017	《信息技术 大数据术语》
3	GB/T 35589-2017	《信息技术 大数据技术参考模型》
4	GB/T 37722-2019	《信息技术 大数据 存储与处理系统功能要求》
5	GB/T 37721-2019	《信息技术 大数据 分析系统功能要求》
6	GB/T 38633-2020	《信息技术 大数据系统运维和管理功能要求》
7	GB/T 38673-2020	《信息技术 大数据 大数据系统基本要求》

8	GB/T 41778-2022	《信息技术 工业大数据 术语》
9	SJ/T 11788-2021	《大数据从业人员能力要求》

（二）专业知识与技术技能

选手需要掌握以下相关知识：

- （1）编程基础：熟悉 Java 或 Python 编程语言。
- （2）数据预处理技术：了解数据采集工具与设备基础知识、数据清洗与处理基础知识等。
- （3）大数据处理技术：需要了解和掌握 Hadoop 等大数据处理技术，以及相关的生态系统和工具。
- （4）分布式计算：大数据处理需要并行分布式计算支持，需要了解和掌握分布式计算框架和分布式计算技术，并能够实现分布式数据计算。
- （5）编程技术：需要熟悉多种编程语言和框架，如 Java、Python 等。
- （6）Web 应用开发技术：需要掌握 Web 应用开发技术 如 HTML、CSS、JavaScript 等，以及相关的 Web 开发框架和库。
- （7）数据可视化：需要了解和掌握数据可视化技术，如 ECharts、Plotly 等，以及相关的可视化工具和库，以便能够将大量的数据展示出来，表达数据信息。

六、技术环境

（一）硬件平台建议配置

电脑	配置信息
CPU	Intel i5 及以上
内存	16GB 及以上
硬盘	500GB 及以上
网络	WI-FI

(二) 软件平台

竞赛原则上采用国产、通用、开源平台。

序号	软件名称	版本	单位	数量
1	Windows10	10+		1
2	Linux Centos	7. x+		3
3	VMware PRO	16. X		1
4	jdk	8. 0. _181-202		3
5	tomcat	9. 0. 58		1
6	maven	3. 5. 4		1
7	Hadoop	2. 6. 0-cdh5. 14. 2		3
11	echarts	4. x+		1
12	eclipse-jee	2021-12-R-win32-x86_64		1

七、奖项设置

1. 参赛选手奖

根据相应组别分别设立一、二、三等奖及优秀奖。其中，一等奖占 10%，二等奖占 20%，三等奖占 30%，剩余为优秀奖。零分卷不得奖。

所有获奖选手均可获得由大赛组委会颁发的获奖证书。

2. 指导教师奖

所有获奖选手的指导教师，均可获得“中国国际数字和软件服务创新大赛优秀指导教师”证书。

3. 参赛学校奖

参赛组织工作成绩突出、经审批符合相关条件的单位，获“中国国际数字和软件服务创新大赛优秀组织单位”称号；

参赛选手成绩优异、经审批符合相关条件的学校，获“中国国际数字和软件服务创新大赛优胜学校”称号。