

上海市高等学校信息技术水平考试

四级大数据与云计算

考试大纲（2022 年版）

一、考试性质

上海市高等学校信息技术水平考试是全市高校统一的教学考试，是检测和评价高校信息技术基础教学水平和教学质量的重要依据之一。该项考试旨在规范和加强高校的信息技术基础教学工作，提高学生的信息技术应用能力。考试对象是高等学校在校学生。考试每年举行一次，通常安排在当年的十月下旬、十一月上旬的星期六或星期日。凡考试成绩达到合格或优秀者，由上海市教育委员会颁发相应的证书。

本考试由上海市教育委员会统一领导，聘请企业和高校专家组成考试委员会，委托上海市教育考试院组织实施。

二、考试目标

上海市高等学校信息技术水平考试（四级）旨在适应新一代信息技术及其应用高速发展的形势，对接现代产业转型升级对信息技术人才的新需要，打通人才培养的需求侧和供给侧，积极引导上海高校开展计算机教学改革，提升大学生信息素养，培养大学生信息技术应用能力，提升大学生的就业竞争力。四级考试面向高年级大学生，由知名信息技术企业支持并参与，考核结果得到相关企业的认可。

四级考试从企业用人标准角度来考核，将企业实际问题作为考试素材，面向实际应用，解决实际问题，充分利用新一代信息技术综合集成，搭建高校与社会企业桥梁，实现知识融合、技能跨界、标准演进。考试依托上海市软件行业协会、上海市知名信息行业企业、高校专家成立命题专家组，以考促教，缩小高校教学与社会需求间的差距，更好地培养大学生的信息素养、计算思维、创新应用和解决实际问题能力，在人工智能、大数据时代更好地提供人才保障。考试根据信息技术发展和行业企业需要区分专业领域进行考核，现设置四个专业领域：人工智能、大数据与云计算、基础软件、网络与信息安全。

考试的目标是考核学生通过运用大数据与云计算技术解决实际应用问题的过程和结果，从而

提高学生对大数据和云计算相关知识、算法、工具、平台、应用的掌握程度，加强学生综合集成与应用能力培养，训练学生的大数据与云计算思维。

考试得到万达信息、帆软软件、星环科技、华为、优刻得、腾讯等企业的支持。

三、考试内容和要求

知识领域	知识单元	知识点	要求
大数据与云计算基础	公共基础	Linux 基础	综合应用
		网络基础	
		服务器基础	
	云计算基础	初识云计算	掌握
		云计算的公共特征与分类	理解
	大数据基础	编程基础	综合应用
数据库基础		综合应用	
深入云计算	云计算原理	虚拟化技术	综合应用
		分布式存储	理解
		分布式计算	理解
		对象存储	理解
		云计算安全	知道
	主机虚拟化	VMware 平台	理解
		OpenStack 框架	综合应用
	容器虚拟化	Docker 容器	综合应用
		Kubernetes 平台	理解
	云原生体系	微服务技术	掌握
		SpringCloud 架构	掌握
		Istio 架构	知道
		DevOps 体系	综合应用
	开发与运维	MQ 消息队列	掌握
		Nginx 与 HaProxy 反向代理	掌握

知识领域	知识单元	知识点	要求
深入云计算	开发与运维	云应用监控	理解
		ELFK 统一日志收集方案	理解
		Serverless 无服务器化技术	知道
大数据平台	大数据基本概念	大数据基础概念	理解
		Hadoop 架构体系	理解
	MapReduce 架构	MapReduce 模型概述	掌握
		Map 和 Reduce 函数	掌握
		MapReduce 工作流程	掌握
		并行计算的实现	理解
	HDFS 分布式文件系统	HDFS 的基本概念	理解
		HDFS 体系结构	知道
		HDFS 命名空间	知道
		HDFS 存储原理	理解
		HDFS 分布式文件系统	掌握
	Zookeeper	Zookeeper 的基本原理和架构	理解
		Zookeeper 的数据模型	知道
	HBase	HBase 的基本原理和架构	理解
		HBase 访问接口	理解
		HBase 数据模型	掌握
		HBase 存储格式和读写数据	综合应用
	Yarn	Yarn 的基本概念和架构	理解
		Yarn 的通信协议	知道
	数据采集	Sqoop 数据同步	掌握
		Flume 日志收集	理解
	离线数据处理	Spark 框架	综合应用
		Hive 数据仓库	理解
	实时数据处理	Flink 框架	理解
		Kafka 分布式消息系统	掌握

知识领域	知识单元	知识点	要求
数据开发与 应用	数据采集	数据采集工具	掌握
		爬虫机制	理解
	数据库深入	高性能数据库	掌握
		非关系数据库	掌握
	数据治理	数据治理概念	理解
		ETL 工具 Kettle 的使用	综合应用
		python 数据清洗	综合应用
	数据仓库与集市	传统的数据仓库	掌握
		实时主动数据仓库	掌握
		数据集成概念	理解
	数据建模与挖掘	数据模型和算法	综合应用
	数据可视化	数据可视化基本概念	理解
		代码级可视化	掌握
		开源工具可视化	综合应用
		商用工具可视化	掌握

备注：

知识与技能的学习考核要求分为**知道**、**理解**、**掌握**和**综合应用**四个层次，其含义分别为：

知道：能识别和记忆相关的学习内容，对相关的知识有初步认识。

理解：初步把握学习内容的由来、作用和使用方法，并能以相应的学习内容为主完成简单的实践。

掌握：以某一学习内容为重点，综合运用其他相关内容，实现给定问题下的实践要求。

综合应用：可以以综合知识去解决实际的应用，完成给定问题合理的解决方案。

四、试卷结构

序号	内容	题型	每题分数	题量	计分
一	基础知识	单选题	1 分	12 题	12 分
二		多选题	2 分	1 题	2 分
三		判断题	2 分	1 题	2 分
四		填空题	2 分	1 题	2 分

序号	内容	题型	每题分数	题量	计分
五	专业方向 (三选一： 深入云计算、大数据 平台、数据 开发与应 用)	单选题	1分	18题	18分
六		多选题	2分	4题	8分
七		判断题	2分	4题	8分
八		填空题	2分	4题	8分
九		简答题	4分	5题	20分
十		综合题(应用)	20分	1题	20分
合计				51题	100分

五、相关说明

1. 考试时间：150分钟。
2. 试卷总分：100分。
3. 等第：不合格、合格、优秀。各等第分数线由考委会划定。
4. 考试方式：考试采用基于网络环境的无纸化上机考试。
5. 考试环境：
 - 上海市高等学校信息技术水平考试通用平台。
 - 操作系统：Windows 10 中文版。
 - 应用开发环境：Python3.6 及其以上版本(包括 sklearn、pandas、numpy 依赖) Excel。
6. 参考教材：
 - [1] 顾炯炯. 云计算架构技术与实践. 清华大学出版社. 2014年.
 - [2] 王伟. 云计算原理与实践. 人民邮电出版社. 2018年.
 - [3] 林子雨. 大数据技术原理与应用. 人民邮电出版社. 2021年.
 - [4] 董西成. 大数据技术体系详解. 机械工业出版社. 2018年.
 - [5] WesMckinney. 利用 Python 进行数据分析. 机械工业出版社. 2018年.

其中，Hadoop 平台、OpenStack 平台、Kettle 工具的使用不限于上述参考教材内容。

