

大数据平台运维 职业技能等级标准

(2020 年 1.0 版)

新华三技术有限公司 制定
2020 年 3 月 发布

目 次

前言.....	3
1 范围.....	4
2 规范性引用文件.....	4
3 术语和定义.....	4
4 适用院校专业.....	9
5 面向职业岗位（群）.....	9
6 职业技能要求.....	9
参考文献.....	25

前 言

本标准按照 GB/T 1.1-2009 给出的规则起草。

本标准起草单位：新华三技术有限公司、工业和信息化部教育与考试中心、无锡职业技术学院、浙江机电职业技术学院、南京信息职业技术学院、长沙民政职业技术学院、重庆电子工程职业学院、贵州轻工职业技术学院。

本标准主要起草人：刘小兵、姚明、肖李晨、于鹏、陈喆、蔡建军、卢涤非、聂明、邓文达、卢建云、汪洪、陈穆衍、白杨、陈永波。

声明：本标准的知识产权归属于新华三技术有限公司，未经新华三技术有限公司同意，不得印刷、销售。

1 范围

本标准规定了大数据平台运维职业技能等级对应的工作领域、工作任务及职业技能要求。

本标准适用于大数据平台运维职业技能培训、考核与评价，相关用人单位的人员聘用、培训与考核可参照使用。

2 规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本标准。凡是不注日期的引用文件，其最新版本适用于本标准。

GB/T 35295-2017 信息技术 大数据 术语

GB/T 5271.1-2000 信息技术 词汇 第1部分：基本术语

高等职业学校专业教学标准（2018年）

本科专业类教学质量国家标准

2019年全国职业院校技能大赛 GZ-2019032 大数据技术与应用赛项规程

大数据安全标准化白皮书（2018版）

大数据标准化白皮书（2018版）

GB/T 37973-2019 《信息安全技术大数据安全管理指南》

GB/T 37722-2019 《信息技术大数据存储与处理系统功能要求》

GB/T 37721-2019 《信息技术大数据分析系统功能要求》

ISO/IEC 20547-4 《信息技术 大数据参考架构 第4部分：安全与隐私保护》

ITU-T Y.3600 《大数据 基于云计算的要求和能力》

3 术语和定义

GB/T 35295-2017、国家、行业标准界定的以及下列术语和定义适用于本标准。

3.1 大数据 Big Data

具有体量巨大、来源多样、生成极快、且多变等特征并且难以用传统数据体系结构有效处理的包含大量数据集的数据。

注：国际上，大数据的 4 个特征普遍不加修饰地直接用 Volume、Variety、Velocity 和 Variability 予以表述，并分别赋予了它们在大数据语境下的定义：

体量 Volume：构成大数据的数据集的规模。

多样性 Variety：数据可能来自多个数据仓库、数据领域或多种数据类型。

速度 Velocity：单位时间的数据流量。

多变性 Variability：大数据其他特征，即体量、速度和多样性等特征都处于多变状态。

3.2 大数据系统 Big Data System

实现大数据参考体系结构的全部或部分功能的系统。

3.3 大数据服务 Big Data Service

基于大数据参考体系结构提供的数据服务。

3.4 集群 Cluster

集群就是一组计算机，它们作为一个整体向用户提供一组网络资源和服务，这些单个的计算机系统就是集群的节点（node）。集群具有可扩展性、高可用性、负载均衡及错误恢复的关键特性。

3.5 虚拟 Virtual

用来修饰一种功能单元，它看起来是实际的，但其功能是通过其他手段得以实现的。

3.6 虚拟机 Virtual Machine, VM（缩写词）

一种虚拟的数据处理系统，它看起来是在某个特定用户的独占使用下，但其功能是通过共享真实数据处理系统的各种资源得以实现的。

3.7 网络功能虚拟化 Network Function Virtualization

对路由器/路由选择、周界防护、远程访问鉴别以及网络流量/载荷监控等网络功能的虚拟应用实现。

注：网络功能虚拟化支持信息系统的高弹性、容错和资源管理，是应对大数据巨大数据体量下用户数据连接的峰、谷起伏问题的至关重要的应用。

3.8 本地虚拟化 Native Virtualization

大数据环境下的一种虚拟化基本形式，按此种形式，在本地裸机上运行管理程序，该程序管理由操作系统和应用组成的多个虚拟机。

3.9 主机虚拟化 Hosted Virtualization

数据环境下的一种虚拟化基本形式，按此种形式，在本地裸机上运行操作系统，在驻留客户操作系统和应用的顶层运行管理程序。

3.10 数据治理 Data Governance

对数据进行处置、格式化和规范化的过程。

注 1：数据治理是数据和数据系统管理的基本要素。

注 2：数据治理及数据全生命周期管理，无论数据是处于静态、动态、未完成状态还是交易状态。

3.11 链接数据 Linked Data

连接其他数据的数据。

3.12 分析 Analytics

根据信息合成知识的过程。

3.13 资源协商 Resource Negotiation

一种支持多租户以及要求高可用性和低延迟的环境的资源访问模式。

注：按此模式，资源管理器是若干节点管理器的集线器；各个客户（或用户）依次请求节点管理器中的应用管理器，紧接前一个请求者的后一个请求者分配到同一个或不同的节点管理器的应用管理器。根据中央处理器（CPU）和存储器可用情况为所请求的任务确定先后次序并在节点提供适当的处理资源。

3.14 集群管理 Cluster Management

在以非关系模型方式驻留数据的集群资源之间提供通信的一种机制。

3.15 数据处理 Data Processing

数据操作的系统执行。

注：术语“数据处理”不能用作“信息处理”的同义词。

3.16 数据管理 Data Management

在数据处理系统中，提供对数据的访问，执行或监视数据的存储，以及控制输入输出操作等功能。

3.17 数据挖掘 Data Mining

从大量的数据中通过算法搜索隐藏于其中信息的过程。

注：一般通过包括统计、在线分析处理、情报检索、机器学习、专家系统（依靠过去的经验法则）和模式识别等方法来实现。

3.18 数据中心 Data Center

由计算机场站（机房）、机房基础设施、信息系统硬件（物理和虚拟资源）、信息系统软件、信息资源（数据）和人员以及相应的规章制度组成的组织。

3.19 数据可视化 Data Visualization

借助图形化手段，清晰有效地传达与沟通信息，是关于数据视觉表现形式的科学技术研究。

3.20 数据平台框架 Data Platform Framework

用于指导实现结合相关应用编程接口（API）访问的逻辑数据组织和分发的集合。

注 1：此类框架一般还包含数据注册和连同语义数据描述（如格式化本体和分类）的元数据服务。逻辑数据组织的覆盖范围从简单限定的平面文件到完全分布式关系数据存储或分栏数据存储。

注 2：这是大数据框架提供者可能提供的一种框架。

3.21 配置 Configuration

信息处理系统中的硬件和软件组织和互连起来的方式。

3.22 下载 To Download

将程序或数据从一个计算机传送到与之相连的资源较少的计算机上，通常是从主计算机传到个人计算机上。

3.23 上载 To Up Load

将程序或数据从一个与之相连的计算机传送到一个资源较多的计算机上，通常是从个人计算机传到主计算机上。

3.24 接口 Interface

两个功能单元共享的边界，它由各种特征（如，功能、物理连接、信号交换等）来定义。

3.25 软件工程 Software Engineering

将科技知识、方法和经验系统地应用到软件的设计、实现、测试和文档编制中，以优化软件的生产、技术支持和质量。

3.26 数据科学 Data Science

根据原始数据，经过整个数据生存周期过程凭借经验合成可用行动的知识的一种科学。

4 适用院校专业

中等职业学校：计算机应用、计算机网络技术、网站建设与管理、软件与信息服务、电子与信息技术、网站安防系统安装与维护等专业。

高等职业学校：计算机应用技术、计算机网络技术、计算机信息管理、计算机系统与维护、软件技术、软件与信息服务、云计算技术与应用、大数据技术与应用、电子商务技术、人工智能技术服务等专业。

应用型本科学校：数据科学与大数据技术、计算机科学与技术、软件工程、网络工程、智能科学与技术、大数据管理与应用、信息管理与信息系统等专业。

5 面向职业岗位（群）

主要面向从事大数据平台安装、配置、规划、部署、实施、优化升级以及大数据平台监控、管理、维护等相关工作的人员。从事大数据平台部署实施，监控、管理、运行维护等相关工作的人员。

6 职业技能要求

6.1 职业技能等级划分

大数据平台运维职业技能等级分为三个等级：初级、中级、高级，三个级别依次递进，高级别涵盖低级别职业技能要求。

【大数据平台运维】（初级）：主要面向大数据平台安装配置、大数据组件安装配置、大数据平台基础实施、大数据平台简单维护及监控工作岗位。从事虚拟化软件安装与使用、基于 Linux 系统的常用服务安装配置、安装配置及运行 Hadoop 集群、安

装配置及运行核心组件、执行客户大数据平台实施方案、监控大数据平台运行状态、查看管理大数据平台日志信息、监控大数据平台服务和资源状态等工作，掌握大数据平台安装和配置方法，理解 Hadoop 核心组件的功能及工作原理，掌握关键组件安装配置方法，理解大数据平台实施流程，熟悉常用集群监控工具的使用方法。

【大数据平台运维】(中级): 主要面向大数据平台高可用性部署实施、大数据组件维护及使用、大数据平台维护及大数据平台优化等工作岗位。从事 Hadoop 高可用集群部署及配置、Hadoop 组件维护及使用、Hadoop 集群节点管理及维护、大数据平台故障诊断及维护等工作，掌握高可用集群 Hadoop 配置方法，熟练使用 shell，掌握 Hadoop 集群节点及其服务的增删改等基本操作方法，熟悉常用系统性能诊断工具及集群监控管理工具，能独立排查和解决大数据平台常见问题，优化集群性能。

【大数据平台运维】(高级): 主要面向大数据平台规划、大数据平台安全管理、大数据平台资源管理、大数据平台优化及升级等工作岗位。从事大数据集群软硬件配置方案拟定、Hadoop 架构方案设计、Hadoop 组件部署方案规划、Hadoop 安全机制规划与实现、大数据平台资源配置及管理、大数据平台优化拓展与升级等工作，熟练运用 shell 进行平台运维，熟练掌握 Hadoop 生态圈组件的工作原理和使用方法，掌握 Hadoop 集群的安全管理机制和方法，熟悉 Hadoop 资源配置和管方法，掌握大数据平台优化策略和方法，熟练 Hadoop 集群软硬件升级方法和操作。

6.2 职业技能等级要求描述

表 1 大数据平台运维职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1.大数据平台安装	1.1 虚拟化软件使用	1.1.1 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立创建虚拟机 1.1.2 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立删除虚拟机

工作领域	工作任务	职业技能要求
		<p>1.1.3 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立修改虚拟机</p> <p>1.1.4 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立访问虚拟机</p>
	1.2 Linux 操作系统安装和使用	<p>1.2.1 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立 Linux 操作系统安装</p> <p>1.2.2 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立区分 Linux 系统种类</p> <p>1.2.3 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立设置 root 密码。</p> <p>1.2.4 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立下载 Linux SSH 服务</p> <p>1.2.5 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立安装 Linux SSH 服务</p> <p>1.2.6 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立安装 Linux JDK</p> <p>1.2.7 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立配置 JDK 环境变量。</p>
	1.3 Hadoop 平台安装	<p>1.3.1 能根据大数据平台安装工作任务要求，按照大数据平台配置指导书，能独立下载 Hadoop 安装包</p> <p>1.3.2 能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立解压 Hadoop 安装包</p> <p>1.3.3 能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立配置 Hadoop 环境变量</p> <p>1.3.4 能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立单机 Hadoop 安装</p> <p>1.3.5 能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立伪分布 Hadoop 安装</p> <p>1.3.6 能根据大数据平台安装工作任务要求，按照大数据平台安装指导书，能独立完全分布 Hadoop 安装</p>
2.大数据平台配置	2.1 主机的基础环境配置	<p>2.1.1 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置主机网络 IP</p> <p>2.1.2 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置主机 DNS</p> <p>2.1.3 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置主机名</p> <p>2.1.4 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能配置集群局域网络连通</p> <p>2.1.5 能根据大数据平台配置工作任务要求，按照大数据平</p>

工作领域	工作任务	职业技能要求
		台配置指导书，能独立生成 SSH 密钥对 2.1.6 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 SSH 认证文件
	2.2 Hadoop 文件参数配置	2.2.1 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 <code>hdfs.site.xml</code> 文件参数 2.2.2 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 <code>core.site.xml</code> 文件参数 2.2.3 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 <code>mapred.site.xml</code> 文件参数 2.2.4 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 <code>yarn.site.xml</code> 文件参数
	2.3 Hadoop 集群运行	2.3.1 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立配置 Hadoop 格式化 2.3.2 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立启动 Hadoop 2.3.3 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立查看 Java 进程 2.3.4 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立查看 HDFS 的报告 2.3.5 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立使用浏览器浏览 Master 节点，查看节点状态 2.3.6 能根据大数据平台配置工作任务要求，按照大数据平台配置指导书，能独立停止 Hadoop
3.大数据平台组件安装配置	3.1 HBase 组件安装配置	3.1.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 HBase 基本知识 3.1.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 HBase 基础环境 3.1.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 HBase 组件 3.1.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 HBase 组件 3.1.5 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 HBase 组件功能
	3.2 Hive 组件安装配置	3.2.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Hive 基本知识 3.2.2 能根据大数据平台组件安装配置工作任务要求，按照

工作领域	工作任务	职业技能要求
		<p>大数据平台组件安装配置指导书，能独立配置 Hive 基础环境</p> <p>3.2.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Hive 组件</p> <p>3.2.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Hive 组件</p> <p>3.2.5 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Hive 组件功能</p>
	3.3 Zookeeper 组件安装 配置	<p>3.3.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Zookeeper 基本知识</p> <p>3.3.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Zookeeper 基础环境</p> <p>3.3.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Zookeeper 组件</p> <p>3.3.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Zookeeper 组件</p> <p>3.3.5 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Zookeeper 组件功能</p>
	3.4 Sqoop 组件安装 配置	<p>3.4.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Sqoop 基本知识</p> <p>3.4.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Sqoop 基础环境</p> <p>3.4.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Sqoop 组件</p> <p>3.4.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Sqoop 组件</p> <p>3.4.5 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Sqoop 组件功能</p>
	3.5 Flume 组件安装 配置	<p>3.5.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能了解 Flume 基本知识</p> <p>3.5.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Flume 基础环</p>

工作领域	工作任务	职业技能要求
		<p>境</p> <p>3.5.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台配置指导书，能独立安装 Flume 组件</p> <p>3.5.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立配置 Flume 组件</p> <p>3.5.5 能根据大数据平台组件安装配置工作任务要求，按照大数据平台组件安装配置指导书，能独立测试 Flume 组件功能</p>
4.大数据平台实施	4.1 大数据平台实施方案的制定	<p>4.1.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能正确理解客户需求</p> <p>4.1.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能编写客户大数据平台实施方案</p> <p>4.1.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能按要求正确执行客户大数据平台实施方案</p> <p>4.1.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能验证客户大数据平台实施方案</p>
	4.2 客户培训方案制定	<p>4.2.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能独立使用文档制作工具</p> <p>4.2.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能制定客户培训方案</p> <p>4.2.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能独立演示操作客户大数据平台</p> <p>4.2.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能培训客户使用大数据平台</p>
	4.3 大数据平台出现的基本问题处理	<p>4.3.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决平台基础环境的基本问题</p> <p>4.3.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决组件配置的基本问题</p> <p>4.3.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决平台启动的基本问题</p> <p>4.3.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台实施指导书，能解决平台运行的基本问题</p>
5.大数据平台监控	5.1 大数据平台运行状态监控	<p>5.1.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控运行状态</p> <p>5.1.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控运行状态</p> <p>5.1.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台运行状态报</p>

工作领域	工作任务	职业技能要求
		表 5.1.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找运行状态异常基本问题
	5.2 大数据平台资源状态监控	5.2.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控资源状态 5.2.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控资源状态 5.2.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台资源状态报表 5.2.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找资源状态异常基本问题
	5.3 大数据平台告警信息监控	5.3.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控告警信息 5.3.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控告警信息 5.3.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台告警信息报表 5.3.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找告警信息异常基本问题
	5.4 大数据平台服务状态监控	5.4.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控服务状态 5.4.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控服务状态 5.4.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台服务状态报表 5.4.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立查找服务状态异常基本问题
	5.5 大数据平台日志信息监控	5.5.1 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过图形监控日志信息 5.5.2 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能通过命令监控日志信息 5.5.3 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能独立制作大数据平台日志信息报表 5.5.4 能根据大数据平台组件安装配置工作任务要求，按照大数据平台监控指导书，能分析基本日志信息

表2 大数据平台运维职业技能等级要求（中级）

工作领域	工作任务	职业技能要求
1. 大数据平台系统架构	1.1 大数据特点及行用场景	<p>1.1.1 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据平台系统架构特点</p> <p>1.1.2 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据平台系统架构原理</p> <p>1.1.3 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据平台系统架构典型企业应用场景</p> <p>1.1.4 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据平台系统架构典型行业应用场景</p>
	1.2 大数据实施和运维流程	<p>1.2.1 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据实施和运维的工作职责</p> <p>1.2.2 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据实施和运维的工作能力素养</p> <p>1.2.3 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据项目实施工作流程</p> <p>1.2.4 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据运维的工作流程</p>
	1.3 大数据发展趋势	<p>1.3.1 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据发展历程</p> <p>1.3.2 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握大数据未来发展趋势</p> <p>1.3.3 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握 Hadoop 生态圈中企业应用典型实战案例</p> <p>1.3.4 能根据大数据平台系统架构工作任务要求,按照大数据平台系统架构指导书,掌握 Hadoop 生态圈中行业应用典型实战案例</p>
2. 大数据平台高可用部署	2.1 Hadoop 集群基础环境配置	<p>2.1.1 能根据大数据平台高可用部署工作任务要求,按照大数据平台高可用部署指导书,能独立配置集群主机之间时钟同步</p> <p>2.1.2 能根据大数据平台高可用部署工作任务要求,按照大数据平台高可用部署指导书,能独立关闭防火墙</p> <p>2.1.3 能根据大数据平台高可用部署工作任务要求,按照大数据平台高可用部署指导书,能独立配置 selinux</p> <p>2.1.4 能根据大数据平台高可用部署工作任务要求,按照大数据</p>

工作领域	工作任务	职业技能要求
		<p>据平台高可用部署指导书，能独立配置 IP 与主机名映射关系</p> <p>2.1.5 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置环境变量</p>
	2.2 Hadoop HA 集群配置	<p>2.2.1 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置高可用 Zookeeper 集群</p> <p>2.2.2 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置高可用集群 Hadoop 文件参数</p> <p>2.2.3 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立配置初始化 JournalNode 服务</p> <p>2.2.4 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 JournalNode 服务</p>
	2.3 Hadoop HA 集群启动	<p>2.3.1 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立初始化 namenode</p> <p>2.3.2 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立复制 namenode 元数据到其它节点</p> <p>2.3.3 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立格式化 zkfs</p> <p>2.3.4 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 Hadoop</p> <p>2.3.5 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 Yarn</p> <p>2.3.6 能根据大数据平台高可用部署工作任务要求，按照大数据平台高可用部署指导书，能独立启动 MapReduce 任务历史服务器</p>
3. 大数据组件维护	3.1 Hbase 组件维护	<p>3.1.1 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立掌握 HBase 组件原理</p> <p>3.1.2 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立卸载 HBase 组件</p> <p>3.1.3 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立创建 HBase 库和表</p> <p>3.1.4 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立在 HBase 库和表中查询、增加、切换、删除操作</p> <p>3.1.5 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立在 HBase 库和表中插入数据</p>

工作领域	工作任务	职业技能要求
		3.1.6 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立退出 HBase 库
	3.2 Hive 组件维护	3.2.1 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立掌握 Hive 组件原理 3.2.2 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立卸载 Hive 组件 3.2.3 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立创建 Hive 库和表 3.2.4 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立在 Hive 库和表中查询、增加、切换、删除操作 3.2.5 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立在 Hive 库和表中插入数据 3.2.6 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立退出 Hive 库
	3.3 Zookeeper 组件维护	3.3.1 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立卸载 Zookeeper 组件 3.3.2 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立修改 Zookeeper 组件 3.3.3 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立配置 Zookeeper 决策选举
	3.4 ETL 组件维护	3.4.1 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立卸载 Sqoop 组件 3.4.2 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立修改 Sqoop 组件 3.4.3 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立使用 Sqoop 进行数据发送和接收 3.4.4 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立下载 Flume 组件 3.4.5 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立修改 Flume 组件
	3.5 Spark 组件维护	3.5.1 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立安装配置 Scala 3.5.2 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立安装 Spark 3.5.3 能根据大数据平台组件维护工作任务要求,按照大数据平台组件维护指导书,能独立修改 Spark 参数 3.5.4 能根据大数据平台组件维护工作任务要求,按照大数据

工作领域	工作任务	职业技能要求
		平台组件维护指导书，能独立启动 Spark 3.5.5 能根据大数据平台组件维护工作任务要求，按照大数据平台组件维护指导书，能独立卸载 Spark
4. 大数据平台优化	4.1 Linux 系统优化	4.1.1 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Linux 系统的内存 4.1.2 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Linux 系统网络 4.1.3 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Linux 系统磁盘 4.1.4 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Linux 文件系统 4.1.5 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Linux 系统缓冲区
	4.2 HDFS 配置优化	4.2.1 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 dfs.replication 文件副本数 4.2.2 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立设置 dfs.block.size 数据块大小 4.2.3 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立设置 mapred.local.dir 磁盘目录 4.2.4 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立设置 dfs.data.dir 优化 IO 读写能力
	4.3 MapReduce 配置优化	4.3.1 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Reduce 节点个数 4.3.2 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Reduce I/O 相关参数 4.3.3 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 reuduce shuffle 阶段并行传输数据的数量 4.3.4 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 tasktracker 并发执行的 reduce 数 4.3.5 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化可并发处理来自 tasktracker 的 RPC 请求数
	4.4 Spark 配置优化	4.4.1 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Spark Streaming 配置 4.4.2 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化 Spark 读取 Kafka 4.4.3 能根据大数据平台优化工作任务要求，按照大数据平台优化指导书，能独立优化读取 Flume

工作领域	工作任务	职业技能要求
		<p>4.4.4 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立优化 Spark 写入 HDFS</p> <p>4.4.5 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立优化 Spark Scala 代码</p>
5. 大数据平台诊断与处理	5.1 Hadoop 及生态圈组件负载均衡诊断与处理	<p>5.1.1 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立对 Hadoop 集群及生态圈组件负载均衡进行诊断</p> <p>5.1.2 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立对高可用 Hadoop 集群及生态圈组件负载均衡进行诊断</p> <p>5.1.3 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立制定 Hadoop 及生态圈组件负载均衡问题解决方案</p> <p>5.1.4 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立处理 Hadoop 及生态圈组件负载均衡问题</p>
	5.2 集群节点故障诊断与处理	<p>5.2.1 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立使用集群日志对节点故障进行诊断</p> <p>5.2.2 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立使用集群告警信息诊断节点故障</p> <p>5.2.3 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立制定集群节点故障解决方案</p> <p>5.2.4 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立处理集群节点故障</p>
	5.3 集群组件服务故障诊断处理	<p>5.3.1 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立使用集群日志诊断组件服务故障问题</p> <p>5.3.2 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立使用集群告警信息诊断组件服务故障问题</p> <p>5.3.3 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立制定集群告警信息诊断组件服务故障问题解决方案</p> <p>5.3.4 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立处理集群告警信息诊断组件服务故障问题</p>

表 3 大数据平台运维职业技能等级要求（高级）

工作领域	工作任务	职业技能要求
1. 大数据平台规划	1.1 Hadoop 集群选型	1.1.1 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能正确选择处理器、内存、存储、网络 1.1.2 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能合理制定集群硬件配置方案 1.1.3 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能独立设计大数据集群网络方案 1.1.4 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能独立设计大数据集群网络高可用方案
	1.2 Hadoop 平台架构设计	1.2.1 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能掌握规划 Hadoop 集群节点高可用方案 1.2.2 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能掌握规划 Hadoop 集群容量方案 1.2.3 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能掌握规划 Hadoop 行业方案 1.2.4 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能掌握规划 Hadoop 企业方案
	1.3 Hadoop 组件部署	1.3.1 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能合理选择 Hadoop 集群组件 1.3.2 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能合理选择 Hadoop 集群组件版本 1.3.3 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能掌握 Hadoop 生态组件的工作原理 1.3.4 能根据大数据平台规划工作任务要求，按照大数据平台规划指导书，能了解 Hadoop 生态圈的发展趋势
2. 大数据平台安全管理	2.1 Hadoop 安全机制实现	2.1.1 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立使用 HDFS 的各类安全机制 2.1.2 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立使用 HDFS 安全组件管理 HDFS 安全 2.1.3 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立使用 MapReduce 的各类安全机制 2.1.4 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立使用 MapReduce 安全组件管理 MapReduce 安全 2.1.5 能根据大数据平台安全管理工作任务要求，按照大数据平台安全管理指导书，能独立使用 Yarn 的各类安全机制

工作领域	工作任务	职业技能要求
		2.1.6 能根据大数据平台安全管理工作任务要求,按照大数据平台安全管理指导书,能独立使用 Yarn 安全组件管理 Yarn 安全
	2.2 大数据平台安全体系实现	2.2.1 能根据大数据平台安全管理工作任务要求,按照大数据平台安全管理指导书,能独立配置大数据平台安全 2.2.2 能根据大数据平台安全管理工作任务要求,按照大数据平台安全管理指导书,能独立配置大数据平台访问安全 2.2.3 能根据大数据平台安全管理工作任务要求,按照大数据平台安全管理指导书,能独立配置大数据平台存储安全 2.2.4 能根据大数据平台安全管理工作任务要求,按照大数据平台安全管理指导书,能独立添加 Hadoop 组件的控制权限 2.2.5 能根据大数据平台安全管理工作任务要求,按照大数据平台安全管理指导书,能独立修改 Hadoop 组件的控制权限
3. 大数据平台资源治理	3.3 大数据平台资源池治理	3.1.1 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立配置静态资源池 3.1.2 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立查看静态资源池状态 3.1.3 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立配置动态资源池 3.1.4 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立使用动态资源池管理资源
	3.1 大数据平台治理体系	3.1.1 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,掌握大数据资产地位确定 3.1.2 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,掌握大数据管理体系机制 3.1.3 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,掌握大数据的数据共享与开放 3.1.4 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,掌握大数据安全与隐私保护
	3.2 大数据平台数据治理	3.2.1 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立对元数据治理 3.2.2 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立对资源池中敏感数据进行标记 3.2.3 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立对数据源与库中的关键字对接

工作领域	工作任务	职业技能要求
		3.2.4 能根据大数据平台资源治理工作任务要求,按照大数据平台资源治理指导书,能独立对数据源与库中的关键字整合
4. 大数据平台优化	4.1 Linux 系统优化	4.1.1 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立避免使用 swap 分区优化 4.1.2 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立调整内存分配策略 4.1.3 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立修改 socket 监听参数 4.1.4 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立优化同时打开文件描述符上限 4.1.5 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立修改 Transparent Huge Pages
	4.2 Hadoop 应用程序优化	4.2.1 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立减少大量小文件输入 4.2.2 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立合理使用分布式缓存 4.2.3 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立合理重用写数据类型 4.2.4 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立对 JVM 缓存调优
	4.3 Hadoop 组件性能优化	4.3.1 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立配置 HDFS 集中缓存管理 4.3.2 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立优化 MapReduce 调度参数 4.3.3 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立优化 Yarn 内存配置 4.3.4 能根据大数据平台优化工作任务要求,按照大数据平台优化指导书,能独立优化 Spark 配置
5. 大数据平台升级	5.1 大数据平台升级	5.1.1 能根据大数据平台升级工作任务要求,按照大数据平台升级指导书,能独立升级大数据平台处理器 5.1.2 能根据大数据平台升级工作任务要求,按照大数据平台升级指导书,能独立升级大数据平台内存 5.1.3 能根据大数据平台升级工作任务要求,按照大数据平台升级指导书,能独立升级大数据平台存储 5.1.4 能根据大数据平台升级工作任务要求,按照大数据平台升级指导书,能独立升级大数据平台网络
	5.2 大数据	5.2.1 能根据大数据平台升级工作任务要求,按照大数据平

工作领域	工作任务	职业技能要求
	平台版本升级	<p>台升级指导书，能独立升级 Hadoop 集群中的 HDFS 配置至高版本</p> <p>5.2.2 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立升级 Hadoop 集群中的 MapReduce 配置至高版本</p> <p>5.2.3 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop 集群运行正常</p> <p>5.2.4 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop HA 集群运行正常</p>
	5.3 Hadoop 组件版本升级	<p>5.3.1 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立升级 Hadoop 组件至高版本</p> <p>5.3.2 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop 组件运行兼容性</p> <p>5.3.3 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop 组件集群运行兼容性</p> <p>5.3.4 能根据大数据平台升级工作任务要求，按照大数据平台升级指导书，能独立验证升级后的 Hadoop 组件之间运行兼容性</p>

参考文献

- [1]国务院关于促进云计算创新发展培育信息产业新业态的意见
- [2]中等职业学校专业目录（征求意见稿）
- [3]普通高等学校高等职业教育（专科）专业目录及专业简介（截至 2018 年）
- [4]普通高等学校本科专业目录（2012 年）
- [5]中等职业学校专业教学标准（试行）
- [6]高等职业学校专业教学标准（2018 年）
- [7]本科专业类教学质量国家标准
- [8]2019 年全国职业院校技能大赛 GZ-2019032 大数据技术与应用赛项规程
- [9]国家职业技能标准编制技术规程（2018 年版）
- [10]中华人民共和国职业分类大典
- [11]战略性新兴产业分类（2018）
- [12]GB/T 4754-2017 国民经济行业分类
- [13]GB/T 5271.1-2000 《信息技术 词汇 第 1 部分：基本术语》
- [14]GB/T 1.1-2009 标准化工作导则
- [15]GB/T 35295-2017 《信息技术 大数据 术语》
- [16]GB/T 35274-2017 《信息安全技术大数据服务安全能力要求》
- [17]GB/T 35589-2017 《信息技术大数据技术参考模型》
- [18]GB/T 36073-2018 《数据管理能力成熟度评估模型》
- [19]大数据安全标准化白皮书（2018 版）
- [20]大数据标准化白皮书（2018 版）

- [21] GB/T 37973-2019 《信息安全技术大数据安全管理指南》
- [22] GB/T 37722-2019 《信息技术大数据存储与处理系统功能要求》
- [23] GB/T 37721-2019 《信息技术大数据分析系统功能要求》
- [24] ISO/IEC 20547-4 《信息技术 大数据参考架构 第 4 部分：安全与隐私保护》
- [25] ITU-T Y. 3600 《大数据 基于云计算的要求和能力》