大数据平台管理与开发

职业技能等级标准

(2021年1.0版)

星环信息科技(上海)股份有限公司 制定
2021年3月 发布

目 次

前言	1
1 范围	2
2 规范性引用文件	2
3 术语和定义	3
4 适用院校专业	8
5 面向职业岗位（群）	8
6 职业技能要求	8
参考文献	20

前 言

本标准按照GB/T 1.1-2020《标准化工作导则 第1部分：标准化文件的结构和起草规则》的规定起草。

本标准起草单位：星环信息科技（上海）股份有限公司、上海立达学院、江西应用技术职业学院、江西现代职业学院、九江职业技术学院、柳州职业技术学院、广西财经学院

本标准主要起草人：周洪明、杨洪山、庞磊、慈雪飞、辛炜、白雪松、高祥兰、魏克俊、杨雪峰、吴圭亮、周学军、刘芑、黄蓉，邵海军、罗胤、胡小春

声明：本标准的知识产权归属于星环信息科技（上海）股份有限公司，未经星环信息科技（上海）股份有限公司同意，不得印刷、销售。

1 范围

本标准规定了大数据平台管理与开发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管理与开发职业技能培训、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

GB/T 35589-2017 信息技术 大数据 技术参考模型

GB/T 35274-2017 信息安全技术 大数据服务安全能力要求

高等职业学校专业教学标准（2018 年）

普通高等学校本科专业类教学质量国家标准（2018 年）

大数据标准化白皮书（2018 版）

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

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 38633-2020 信息技术 大数据系统运维和管理功能要求

GB/T 38643-2020 信息技术 大数据分析系统功能测试要求

GB/T 38676-2020 信息技术 大数据存储与处理系统功能测试要求

ITU-T F.743.20 Assessment framework for big data infrastructure

(大数据基础设施评测框架)

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：

- a) 体量 Volume：构成大数据的数据集的规模。
- b) 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。
- c) 速度 Velocity：单位时间的数据流量。
- d) 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。

[GB/T 35295-2017, 定义 2.1.1]

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。

[GB/T 35295-2017, 定义 2.1.14]

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。

[GB/T 35295-2017, 定义 2.1.15]

3.4 数据 data

信息的可再解释的形式化表示，以适用于通信、解释或处理。

[GB/T 35295-2017, 定义 2.2.1]

3.5 数据处理 data processing

数据操作的系统执行。

[GB/T 35295-2017, 定义 2.2.2]

3.6 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。

[GB/T 35295-2017, 定义 2.2.3]

3.7 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。

[GB/T 35295-2017, 定义 2.1.59]

3.8 分布式计算 distributed computing

一种覆盖存储层和处理层的、用于实现多类型程序设计算法模型的计算模式。

[GB/T 35295-2017, 定义 2.1.22]

3.9 分布式数据处理 distributed data processing; DDP

将操作分散到计算机网络的各结点进行的数据处理。

[GB/T 35295-2017, 定义 2.2.6]

3.10 结构化数据 structured data

一种数据表示形式，按此种形式，由数据元素汇集而成的每个记录的结构都

是一致的并且可以使用关系模型予以有效描述。

[GB/T 35295-2017, 定义 2.2.13]

3.11 非结构化数据 unstructured data

不具有预定义模型或未以预订方式组织的数据。

[GB/T 35295-2017, 定义 2.1.25]

3.12 数据仓库 data warehouse

在数据准备之后用于永久性存储数据的数据库。

[GB/T 35295-2017, 定义 2.1.35]

3.13 数据治理 data governance

对数据进行处置、格式化和规范化的过程。

[GB/T 35295-2017, 定义 2.1.43]

3.14 数据集 data set

数据记录汇聚的数据形式。

[GB/T 35295-2017, 定义 2.1.46]

3.15 关系型数据库 relational database

数据按关系模型来组织的数据库。

[GB/T 35295-2017, 定义 2.2.5]

3.16 元数据 metadata

关于数据或数据元素的数据（可能包括其数据描述），以及关于数据拥有权、存取路径、访问权和数据易变性的数据。

[GB/T 35295-2017, 定义 2.2.7]

3.17 流数据 streaming data

经由接口传递，从连续运行的数据源产生的数据。

[GB/T 35295-2017, 定义 2.1.24]

3.18 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。

[GB/T 35295-2017, 定义 2.2.9]

3.19 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。

[GB/T 35295-2017, 定义 2.2.12]

3.20 基础设施框架 infrastructure framework

由网络、计算、存储和环境等功能构件构成的一种集合。

注 1：网络、计算、存储和环境的解释如下：

a) 网络：支持将数据从一个资源传输到另一个资源的资源（如，已定义的物理资源、软件资源、虚拟资源等）。

b) 计算：执行和驻留其他大数据系统构件（如，物理资源、操作系统、虚拟现实、逻辑分布）的软件的物理处理器和存储器。

c) 存储：在大数据系统中保存数据的资源（如，存储器、本地磁盘、独立磁盘的软/硬件冗余阵列、存储域网、附网存储）。

d) 环境：建立大数据系统时必须考虑的物理辅助资源（如，供电、冷却等）。

注 2：这是大数据框架提供者可能提供的一种框架。

[GB/T 35295-2017, 定义 2.2.9]

3.21 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。

[GB/T 35295-2017, 定义 2.2.10]

3.22 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。

[GB/T 35295-2017, 定义 2.1.40]

3.23 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。

[GB/T 35295-2017, 定义 2.1.63]

3.24 主机虚拟化 hosted 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。

[GB/T 35295-2017, 定义 2.1.64]

3.25 容器式虚拟化 containerized virtualization

大数据环境下的一种虚拟化基本形式，按此种形式，在本地裸机上运行操作

系统，管理程序功能嵌入操作系统中，应用在内部各个容器中运行，控制或限制对操作系统和物理机器资源的访问。

[GB/T 35295-2017, 定义 2.1.65]

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、软件与信息服务、电子与信息技术、计算机信息管理、多媒体技术、大数据技术、物联网技术应用等专业。

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、人工智能技术服务、物联网工程技术等专业。

应用型本科学校：数据科学与大数据技术、大数据管理与应用、智能科学与技术、人工智能、计算机科学与技术、软件工程、物联网工程、网络工程、信息工程等专业。

5 面向职业岗位（群）

主要面向互联网、金融、制造、能源、医疗、交通、电力、通信、教育等行业的信息技术服务领域，从事大数据架构设计、大数据开发、大数据分析处理、大数据平台维护管理等工作岗位的相关人员。

6 职业技能要求

6.1 职业技能等级划分

大数据平台管理与开发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。

【大数据平台管理与开发】(初级)：主要面向初级大数据开发、初级大数据采集开发、初级大数据平台运维管理等工作岗位。初步掌握大数据核心思想、以

及大数据技术生态圈中的核心技术，能够按照项目要求，独立完成简单的开发、组件运维、平台监控管理等工作任务。

【大数据平台管理与开发】(中级): 主要面向中级大数据开发、中高级大数据采集开发、大数据分析处理、大数据平台维护管理等工作岗位。基本掌握大数据各项核心技术组件，并对各个组件的使用与开发有一定的积累，能够胜任大数据基础开发、大数据分析处理、大数据平台管理等工作任务。

【大数据平台管理与开发】(高级): 主要面向高级大数据开发、高级大数据平台管理、大数据项目架构等工作岗位。在熟练掌握各项大数据技术组件的前提下，熟悉各类典型的大数据项目应用场景，能够根据客户业务需求及项目建设方案，完成项目的模块开发、大数据组件故障维护、大数据平台综合管理等工作任务。

6.2 职业技能等级要求描述

表 1 大数据平台管理与开发职业技能等级要求（初级）

工作领域	工作任务	职业能力要求
1. 大数据平台集群硬件管理	1.1 大数据平台服务器基础管理	1.1.1 按照大数据平台管理指导书，能够独立登录 Linux 服务器； 1.1.2 按照大数据平台管理指导书，能够独立完成 Linux 文件目录的创建、查看、删除等操作； 1.1.3 按照大数据平台管理指导书，能够独立完成 Linux 文件的创建、查看、移动、复制、删除等操作； 1.1.4 按照大数据平台管理指导书，能够独立掌握 Linux vi/vim 文本编辑操作。
	1.2 大数据平台集群管理	1.2.1 按照大数据平台管理指导书，能够通过集群图形化管理应用程序启动集群； 1.2.2 按照大数据平台管理指导书，能够通过集群图形化管理应用程序停止集群； 1.2.3 按照大数据平台管理指导书，能够独

工作领域	工作任务	职业能力要求
		立通过集群图形化管理应用程序查看集群健康状态; 1. 2. 4 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看集群历史警报情况。
	1. 3 大数据平台节点管理	1. 3. 1 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序添加新的节点; 1. 3. 2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序删除节点; 1. 3. 3 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序启动节点所有角色; 1. 3. 4 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序停止节点所有角色; 1. 3. 5 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看节点运行状态。
	1. 4 大数据平台资源监控管理	1. 4. 1 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看集群资源使用情况; 1. 4. 2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看节点资源使用情况。
2. 大数据平台组件可视化监控管理	2. 1 HDFS 组件运行可视化监控管理	2. 1. 1 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序启动 HDFS 组件; 2. 1. 2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序停止 HDFS 组件; 2. 1. 3 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看 HDFS 组件的运行健康状态。
	2. 2 Yarn 组件运行可视化监控管理	2. 2. 1 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序启动 Yarn 组件; 2. 2. 2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序停止 Yarn 组件; 2. 2. 3 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看 Yarn

工作领域	工作任务	职业能力要求
		组件的运行健康状态。
	2.3 Zookeeper 组件运行可视化监控管理	2.3.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 Zookeeper 组件； 2.3.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 Zookeeper 组件； 2.3.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 Zookeeper 组件的运行健康状态。
	2.4 Hive 组件运行可视化监控管理	2.4.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 Hive 组件； 2.4.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 Hive 组件； 2.4.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 Hive 组件的运行健康状态。
	2.5 HBase 组件运行可视化监控管理	2.5.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 HBase 组件； 2.5.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 HBase 组件； 2.5.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看 HBase 组件的运行健康状态。
3. 大数据组件使用	3.1 SQL 基本操作命令使用	3.1.1 按照大数据平台开发指导书，能够独立完成数据库的创建、查看、删除等操作； 3.1.2 按照大数据平台开发指导书，能够独立完成表的创建、查看、删除等操作； 3.1.3 按照大数据平台开发指导书，能够独立完成数据的写入、更新、删除等操作； 3.1.4 按照大数据平台开发指导书，能够独立完成表数据的简单查询； 3.1.5 按照大数据平台开发指导书，能够独立完成表数据的条件过滤查询； 3.1.6 按照大数据平台开发指导书，能够独立完成表数据的子查询。
	3.2 HDFS 组件使用	3.2.1 理解 HDFS 存储机制； 3.2.2 按照大数据平台开发指导书，能够独

工作领域	工作任务	职业能力要求
		立通过 Shell 命令的方式完成 HDFS 目录的创建、删除、修改等; 3.2.3 按照大数据平台开发指导书,能够独立通过 Shell 命令的方式完成 HDFS 文件的上传、查看、下载等; 3.2.4 按照大数据平台开发指导书,能够独立通过 Shell 命令的方式完成 HDFS 文件及目录权限的查看、授予、删除等。
	3.3 Yarn 组件使用	3.3.1 理解 Yarn 的基础原理; 3.3.2 了解 Yarn 的三种任务调度机制; 3.3.3 按照大数据平台开发指导书,能够独立通过 Shell 命令的方式完成 Yarn 中任务的提交、查看、终止等。
	3.4 Hive 组件使用	3.4.1 理解 Hive 基础原理; 3.4.2 按照大数据平台开发指导书,能够独立完成 Hive 数据库的创建、查看、删除等操作; 3.4.3 按照大数据平台开发指导书,能够独立完成 Hive 内部表的创建、查看、删除等操作; 3.4.4 按照大数据平台开发指导书,能够独立完成 Hive 外部表的创建、查看、删除等操作; 3.4.5 按照大数据平台开发指导书,能够独立完成 Hive 表数据的写入操作; 3.4.6 按照大数据平台开发指导书,能够独立完成 Hive 表的简单查询、过滤查询等操作。
	3.5 HBase 组件使用	3.5.1 理解 HBase 存储机制; 3.5.2 按照大数据平台开发指导书,能够独立通过 Shell 命令完成 HBase 命名空间的创建、查看、删除等操作; 3.5.3 按照大数据平台开发指导书,能够独立通过 Shell 命令完成 HBase 表的创建、查看、禁用、删除等操作; 3.5.4 按照大数据平台开发指导书,能够独立通过 Shell 命令完成 HBase 表数据的写入操作; 3.5.5 按照大数据平台开发指导书,能够独立通过 Shell 命令完成 HBase 表数据的全表查询、过滤查询等操作。
4. 大数据 ETL	4.1 Sqoop 工具	4.1.1 熟悉 Sqoop 工具使用规范;

工作领域	工作任务	职业能力要求
工具使用	使用	4.1.2 按照大数据平台开发指导书，能够独立编写 Sqoop 脚本完成查看数据库、查看数据表等操作； 4.1.3 按照大数据平台开发指导书，能够独立编写 Sqoop 脚本完成全量数据导入操作； 4.1.4 按照大数据平台开发指导书，能够独立编写 Sqoop 脚本完成全量数据导出操作。
	4.2 Flume 工具使用	4.2.1 熟悉 Flume 工具使用规范； 4.2.2 按照大数据平台开发指导书，能够基本掌握 Flume 中 Exec Source、Avro Source 等的使用方法； 4.2.3 按照大数据平台开发指导书，能够基本掌握 Flume 中 Memory Channe 的基础使用方法； 4.2.4 按照大数据平台开发指导书，能够基本掌握 FLume 中 Logger Sink、Avro Sink 等的使用方法。
	4.3 Kafka 工具使用	4.3.1 熟悉 Kafka 工作原理； 4.3.2 按照大数据平台开发指导书，能够基本掌握 Kafka Shell 命令用法。

表 2 大数据平台管理与开发职业技能等级要求（中级）

工作领域	工作任务	职业能力要求
1. 大数据平台组件维护管理	1.1 大数据平台服务器管理	1.1.1 按照大数据平台管理指导书，能够独立完成 Linux 文本处理的工作； 1.1.2 按照大数据平台管理指导书，能够独立完成 Linux 用户、权限管理工作； 1.1.3 按照大数据平台管理指导书，能够独立完成 Linux 硬件资源监控管理的工作； 1.1.4 按照大数据平台管理指导书，能够独立完成 Linux 软件安装的工作。
	1.2 Hadoop 组件维护管理	1.2.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序启动 Hadoop 各组件； 1.2.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序停止 Hadoop 各组件； 1.2.3 按照大数据平台管理指导书，能够独

工作领域	工作任务	职业能力要求
		立通过集群图形化管理应用程序查看 Hadoop 各组件的配置; 1. 2. 4 根据实际大数据平台管理要求, 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序修改 Hadoop 各组件的配置。
	1. 3 Hive 组件维护管理	1. 3. 1 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序启动 Hive 组件; 1. 3. 2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序停止 Hive 组件; 1. 3. 3 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看 Hive 组件的配置; 1. 3. 4 根据实际大数据平台管理要求, 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序修改 Hive 组件的配置。
	1. 4 HBase 组件维护管理	1. 4. 1 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序启动 HBase 组件; 1. 4. 2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序停止 HBase 组件; 1. 4. 3 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序查看 HBase 组件的配置; 1. 4. 4 根据实际大数据平台管理要求, 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序修改 HBase 组件的配置。
2. 大数据组件开发	2. 1 SQL 数据分析开发	2. 1. 1 按照大数据平台开发指导书, 能够独立完成数据库、表的权限管理操作; 2. 1. 2 按照大数据平台开发指导书, 能够独立完成表数据的 Join 连接查询操作; 2. 1. 3 按照大数据平台开发指导书, 能够独立完成表的索引创建、使用等操作; 2. 1. 4 按照大数据平台开发指导书, 能够独立完成视图的创建、更新、删除等操作; 2. 1. 5 按照大数据平台开发指导书, 能够独立掌握函数的使用。
	2. 2 MapReduce	2. 2. 1 理解 MapReduce 编程模型、运行机制;

工作领域	工作任务	职业能力要求
	组件开发	2.2.2 按照大数据平台开发指导书，能够理解基础的 MapReduce 统计分析案例； 2.2.3 按照大数据平台开发指导书，能够独立提交 MapReduce 计算任务； 2.2.4 按照大数据平台开发指导书，能够独立查看 MapReduce 计算任务执行结果。
	2.3 Hive 组件开发	2.3.1 理解 Hive 架构原理； 2.3.2 按照大数据平台开发指导书，能够掌握 Hive 复杂数据类型的开发； 2.3.3 按照大数据平台开发指导书，能够掌握 Hive 复杂查询的开发； 2.3.4 按照大数据平台开发指导书，能够掌握 Hive 函数的使用。
	2.4 HBase 组件开发	2.4.1 理解 HBase 架构原理； 2.4.2 按照大数据平台开发指导书，能够以 SQL 方式完成表的创建、查看、禁用、删除等开发任务； 2.4.3 按照大数据平台开发指导书，能够以 SQL 方式完成表数据写入的开发任务； 2.4.4 按照大数据平台开发指导书，能够以 SQL 方式完成表数据查询的开发任务。
	2.5 Spark 开发	2.5.1 理解 Spark 基本原理； 2.5.2 按照大数据平台开发指导书，能够掌握 Spark 开发规范； 2.5.3 按照大数据平台开发指导书，能够完成 RDD 的基础开发； 2.5.4 按照大数据平台开发指导书，能够独立提交 Spark 计算任务； 2.5.5 按照大数据平台开发指导书，能够独立提交 Spark 计算任务执行结果。
3. 大数据 ETL 工具开发	3.1 Sqoop 工具开发	3.1.1 按照大数据平台开发指导书，能够使用 Sqoop 完成关系型数据库数据的采集工作； 3.1.2 按照大数据平台开发指导书，能够使用 Sqoop 完成大数据平台数据的导出工作； 3.1.3 按照大数据平台开发指导书，能够掌握 Sqoop 全量数据处理、以及增量数据处理； 3.1.4 根据实际大数据数据采集的工作要求，按照大数据平台开发指导书，能够利用 Sqoop 完成关系型数据库数据采集工作。
	3.2 Flume 工具开发	3.2.1 按照大数据平台开发指导书，能够熟练掌握 Flume 中 Spooling Directory

工作领域	工作任务	职业能力要求
		Source、Kafka Source 等的开发; 3.2.2 按照大数据平台开发指导书, 能够熟练掌握 Flume 中 Memory Channel 的高级配置用法; 3.2.3 按照大数据平台开发指导书, 能够熟练掌握 FLume 中 HDFS Sink、Kafka Sink 等的开发; 3.2.4 根据实际大数据数据采集的工作要求, 按照大数据平台开发指导书, 能够利用 Flume 完成日志数据采集工作。
	3.3 Kafka 工具开发	3.3.1 熟练掌握 Kafka 基础原理; 3.3.2 按照大数据平台开发指导书, 能够熟练掌握 Kafka 的 Productor、Consumer 等操作命令; 3.3.3 根据实际大数据实时数据采集的工作要求, 按照大数据平台开发指导书, 能够利用 Kafka 完成实时数据采集工作。

表 3 大数据平台管理与开发职业技能等级要求（高级）

工作领域	工作任务	职业能力要求
1. 大数据平台安全管控及故障维护管理	1.1 大数据平台账户管理	1.1.1 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序熟练使用大数据平台管理员账号; 1.1.2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序完成平台内用户的新增、查询、修改、删除等工作; 1.1.3 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序完成平台内角色的新增、查询、修改、删除等工作; 1.1.4 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序完成平台内组的新增、查询、修改、删除等工作; 1.1.5 根据实际大数据平台账户管理工作的要求, 独立完成用户、角色、组的配置工作。
	1.2 大数据平台集群安全认证	1.2.1 理解 Kerberos 基本概念; 1.2.2 按照大数据平台管理指导书, 能够独立通过集群图形化管理应用程序完成平台内用户权限的授予、查询、删除等工作;

工作领域	工作任务	职业能力要求
		<p>1.2.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序完成平台内角色权限的授予、查询、删除等工作；</p> <p>1.2.4 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序完成平台内组权限的授予、查询、删除等工作；</p> <p>1.2.5 根据实际大数据平台安全管理工作的要求，独立完成用户、角色、组的安全权限配置工作。</p>
	1.3 大数据平台组件权限管控	<p>1.3.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序完成各大大数据平台组件安全管控的开启、停止、配置等工作；</p> <p>1.3.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序完成各大大数据平台组件的授权、配额管理等工作。</p>
	1.4 大数据平台组件故障维护管理	<p>1.4.1 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看集群、节点的CUP、内存、磁盘、网络的实时使用情况；</p> <p>1.4.2 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看集群、节点的CUP、内存、磁盘、网络的历史使用情况；</p> <p>1.4.3 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看各组件的资源概要信息、服务历史信息、节点信息、统计信息等；</p> <p>1.4.4 按照大数据平台管理指导书，能够独立通过集群图形化管理应用程序查看集群警报通知；</p> <p>1.4.5 根据实际大数据平台管理运行情况，按照大数据平台管理指导书，能够排查、解决各组件的运行故障。</p>
2. 大数据数据仓库场景开发	2.2 数据仓库建设方案设计	<p>2.2.1 熟练掌握数据仓库相关大数据技术组件；</p> <p>2.2.2 熟练掌握典型的数据仓库建设方案；</p> <p>2.2.3 能够依据需求说明书，按照大数据平台开发指导书，独立设计建设方案。</p>
	2.2 数据仓库模型设计	<p>2.3.1 熟练掌握数据仓库星型模型、雪花模型；</p> <p>2.3.2 能够依据需求说明书，按照大数据平</p>

工作领域	工作任务	职业能力要求
		台开发指导书，独立设计合理的数据仓库模。
	2.3 数据 ETL 处理	2.4.1 熟练掌握大数据 ETL 数据采集、清洗、加载的工具开发使用； 2.4.2 按照大数据平台开发指导书，能够独立选择合理的工具； 2.4.3 能够依据建设方案，按照大数据平台开发指导书，独立完成 ETL 开发的工作。
	2.4 数据分析	2.5.1 熟练掌握大数据常用的数据分析技术； 2.5.2 根据建设方案，按照大数据平台开发指导书，选择合理的数据分析工具； 2.5.3 能够依据建设方案，按照大数据平台开发指导书，独立完成数据分析开发的工作。
3. 大数据实时流处理场景开发	3.1 实时流处理项目需求分析	3.1.1 能够独立与客户进行合理的需求沟通； 3.1.2 能够独立为客户执行详细的需求说明书； 3.1.3 掌握金融、政府、互联网、交通、能源等行业的典型大数据实时流处理项目案例。
	3.2 实时流处理建设方案设计	3.2.1 熟练掌握实时流处理相关大数据技术组件； 3.2.2 熟练掌握典型的实时流处理建设方案； 3.2.3 能够按照大数据平台开发指导书，独立选择合理的技术方案； 3.2.4 能够依据需求说明书，按照大数据平台开发指导书，独立设计建设方案。
	3.3 实时数据采集	3.3.1 熟练掌握典型的实时数据采集工具； 3.3.2 能够依据建设方案，按照大数据平台开发指导书，独立完成实时数据采集的开发工作。
	3.4 实时流处理分析	3.4.1 熟练掌握典型的实时数据分析技术； 3.4.2 能够依据建设方案，按照大数据平台开发指导书，独立完成实时流处理分析的开发工作。
4. 大数据综合搜索场景开发	4.1 综合搜索项目需求分析	4.1.1 能够独立与客户进行合理的需求沟通； 4.1.2 能够独立为客户执行详细的需求说明书；

工作领域	工作任务	职业能力要求
		4.1.3 掌握金融、政府、互联网、交通、能源等行业的典型大数据综合搜索项目案例。
	4.2 综合搜索建设方案设计	4.2.1 熟练掌握综合搜索相关大数据技术组件； 4.2.2 熟练掌握典型的综合搜索建设方案； 4.2.3 能够依据需求说明书，按照大数据平台开发指导书，独立设计建设方案。
	4.3 综合搜索开发	4.3.1 熟练掌握大数据全文搜索技术； 4.3.2 熟练掌握大数据高并发检索技术； 4.3.3 能够依据建设方案，按照大数据平台开发指导书，独立完成综合搜索的开发工作。

参考文献

- [1] GB/T 35295-2017 信息技术 大数据 术语
- [2] GB/T 5271.1-2000 信息技术 词汇 第1部分:基本术语
- [3] GB/T 35589-2017 信息技术 大数据 技术参考模型
- [4] GB/T 35274-2017 信息安全技术 大数据服务安全能力要求
- [5] 中等职业学校专业目录（2020年）
- [6] 普通高等学校高等职业教育（专科）专业目录（2020年）
- [7] 普通高等学校本科专业目录（2020年版）
- [8] 中等职业学校专业教学标准（试行）
- [8] 高等职业学校专业教学标准（2018年）
- [9] 普通高等学校本科专业类教学质量国家标准（2018年）
- [10] 大数据标准化白皮书（2018版）
- [11] 中华人民共和国职业分类大典
- [12] 国务院关于印发国家职业教育改革实施方案的通知（国发〔2019〕4号）
- [13] 国务院关于印发促进大数据发展行动纲要的通知（国发〔2015〕50号）
- [14] 国家职业技能标准编制技术规程（2018年版）
- [15] 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
- [16] GB/T 37973-2019 信息安全 技术大数据安全管理指南
- [17] GB/T 37722-2019 信息技术大数据存储与处理系统功能要求
- [18] GB/T 38633-2020 信息技术大数据系统运维和管理功能要求

[19] GB/T 38643-2020 信息技术大数据分析系统功能测试要求

[20] GB/T 38676-2020 信息技术大数据存储与处理系统功能测试要求

[21] ITU-T F.743.20 Assessment framework for big data infrastructure (大数据基础设施评测框架)