

工业大数据应用与开发 职业技能等级标准

(2021年2.1版)

沈阳新松机器人自动化股份有限公司 制定
2021年4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.....	2
4 适用院校专业.....	3
5 面向职业岗位（群））	4
6 职业技能要求.....	4
参考文献.....	11

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》给出的规则起草。

本标准起草单位：由沈阳新松机器人自动化股份有限公司主持，联合沈阳中德新松教育科技集团有限公司、新松机器人教育科技（苏州）有限公司、漳州弈恩现实信息科技有限公司、中国科学院自动化研究所、中国科学院计算技术研究所等单位共同制定。

本标准主要起草人：范存艳、李劲松、史泽林、刘班、靳玉峰、王利臣、杨静、刘松、韩海峰、许昊鹏、夏文通。

声明：本标准的知识产权归属于沈阳新松机器人自动化股份有限公司，未经沈阳新松机器人自动化股份有限公司同意，不得印刷、销售。

1 范围

本标准规定了工业大数据应用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。

本标准适用于工业大数据应用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 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

GB/T 33745-2017 物联网 术语

GB/T 5271.1-2000 信息技术 词汇 第1部分：基本术语

GB/T 38619-2020 工业物联网 数据采集结构化描述规范

GB/T 34068-2017 物联网总体技术 智能传感器接口规范

GB/T 32919-2016 信息安全技术 工业控制系统安全控制应用指南

SJ/T 11666.1—2016 制造执行系统（MES）规范 第1部分：模型和术语

3 术语和定义

国家标准、行业标准界定的以及下列术语适用于本标准。

3.1 大数据 big data

具有体量巨大、来源多样、生成极快、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。

[GB/T 35295-2017 信息技术 大数据 术语]

3.2 大数据应用 big data application

运用编程手段或非编程手段完成的大数据系统（big data system）的全部或部分功能。

[GB/T 35295-2017 信息技术 大数据 术语]

3.3 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 信息技术 大数据 术语]

3.4 数据采集 data acquisition

数据采集又称数据获取，是利用一种装置从系统外部采集数据并输入到系统内部的一个接口。

[GB/T 35295-2017 信息技术 大数据 术语]

3.5 数据清洗 Data cleaning

数据清洗是指在数据分析前对数据进行的一些处理，主要是清理异常值、纠正错误数据。

[GB/T 35295-2017 信息技术 大数据 术语]

3.6 数据处理 data processing

数据操作的系统执行。

[GB/T 35295-2017 信息技术 大数据 术语]

3.7 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 信息技术 大数据 术语]

3.8 数据可视化 data visualization

数据可视化旨在借助于图形化手段，清晰有效地传达与沟通信息。

[GB/T 35295-2017 信息技术 大数据 术语]

3.9 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 信息技术 大数据 术语]

3.10 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 信息技术 大数据 术语]

3.11 物联网 internet of things; IOT

通过感知设备，按照约定协议连接物、人、系统和信息资源，实现对物理和虚拟世界的信息进行处理并做出反应的智能服务系统²。

[GB/T 34068-2017 物联网总体技术 智能传感器接口规范]

3.12 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 信息技术 大数据 术语]

3.13 制造执行系统 (MES) Manufacturing Execution System (MES)

通过信息传递，对从订单下达到产品完成的整个生产过程进行优化管理。

[SJ/T 11666.1—2016 制造执行系统 (MES) 规范 第1部分：模型和术语]

注：当工厂发生实时事件时，MES能够及时作出反应、报告，并根据当前的准确数据进行指导和处理。制造执行系统还通过双向的直接通信在企业内部和整个产品供应链中提供有关产品行为的关键任务信息。

4 适用院校专业

中等职业学校：软件与信息服务、计算机应用、计算机网络技术、物联网技术应用、电子技术应用、通信技术、移动应用技术与服务等专业。

高等职业学校：计算机应用技术、软件技术、云计算技术与应用、大数据技术与应用、人工智能技术服务、计算机信息管理、软件与信息服务、电子信息工程技术、物联网应用技术、计算机网络技术等专业。

本科学校：软件工程、数据科学与大数据技术、人工智能、计算机科学与技术、电子信息工程、电子科学与技术、物联网工程、信息管理与信息系统、信息工程等专业。

5 面向职业岗位（群）

工业大数据应用与开发（初级）：主要面向工业企业的大数据平台初级运维工程师、初级数据库管理工程师、大数据初级ETL工程师、数据可视化初级工程师等岗位。

工业大数据应用与开发（中级）：主要面向工业企业的数据库管理工程师、大数据平台运维工程师、数据可视化工程师、大数据ETL工程师、大数据开发工程师等岗位。

工业大数据应用与开发（高级）：主要面向工业企业的大数据平台运维工程师、高级数据库管理工程师、高级大数据ETL工程师、高级数据可视化工程师、高级大数据开发工程师、数据分析师、数据挖掘工程师、项目经理等岗位。

6 职业技能要求

6.1 职业技能等级划分

工业大数据应用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。

【工业大数据应用与开发】（初级）：根据工业企业的业务需求，完成工业大数据软件的安装部署，能够监控集群的运行状态，进行简单的运维处理；能够进行简单的数据库管理、数据查询及处理操作；能够使用工具完成简单的数据采集与数据清洗；能够进行Hadoop分布式文件系统HDFS的存储；能够使用工具完成简单的数据可视化。

【工业大数据应用与开发】（中级）：根据工业企业的业务需求，能够实现工业大数据应用编程，能够使用编程方式实现数据采集、分布式数据库Hbase应用、数据仓库hive应用、内存数据库redis应用，能够使用工具实现复杂功能的数据可视化。

【工业大数据应用与开发】（高级）：根据工业企业的业务需求，能够对实时数据使用流式计算技术进行处理；能够使用spark内存计算技术对批量数据进行处理；能够使用编程技术完成数据清洗、数据分析、数据挖掘、数据可视化。

6.2 职业技能等级要求描述

表 1 工业大数据应用与开发职业技能等级要求（初级）

工作领域	工作任务	职业技能要求
1. 工业大数据软件安装与维护	1.1 工业大数据软件安装部署	1.1.1 能够根据工业大数据业务场景的需求，选择合适的工业大数据软件，并根据运行要求，选择合适的软硬件环境； 1.1.2 能够独立安装 Linux 操作系统； 1.1.3 能够根据给定的系统软硬件环境，独立完成工业大数据软件的安装与部署； 1.1.4 能够独立完成软件安装后的测试。
	1.2 工业大数据软件维护	1.2.1 能够使用工具监控工业大数据软件的运行状况； 1.2.2 能够根据工业大数据平台软件的运行状况，独立完成软硬件资源的调整。
2. 关系型数据库应用开发	2.1 关系型数据库部署与管理	2.1.1 能够根据工业大数据业务场景的需求，选择合适的操作系统，独立安装 mysql 数据库； 2.1.2 能够使用适合的客户端软件远程访问数据库服务器； 2.1.3 能够根据业务需求，选择合适的存储引擎并创建数据库； 2.1.4 能够根据业务需求，创建用户并分配用户适当的权限。
	2.2 关系型数据库的数据查询	2.2.1 能够根据业务需求，完成单表数据的条件查询和排序操作； 2.2.2 能够根据业务需求，完成多表连接查询； 2.2.3 能够根据业务需求，使用分组函数完成分组统计功能； 2.2.4 能够根据业务需求，使用子查询实现高级查询功能； 2.2.5 能够根据业务需求，使用常用函数完成复杂的查询功能。
	2.3 数据操作与事务控制	2.3.1 能够根据业务需求，进行业务数据的增加、修改、删除操作； 2.3.2 能够根据业务需求，完成事务的控制，保证数据的一致性；
	2.4 数据对象管理	2.4.1 能够根据业务需求，完成普通表的维护，并设置有效地数据约束； 2.4.2 能够根据业务需求，选择合适的分区表，实现大数据的分区管理； 2.4.3 能够根据业务需求，实现视图、索引对象的管理。

3. 工业大数据存储与运维	3.1 分布式文件系统 HDFS 的存储及运维	3.1.1 能够完成分布式文件系统的数据存储； 3.1.2. 能够完成 Hadoop 分布式存储的运维。
4. 工业数据采集与清洗	4.1 工业数据采集	4.1.1 能够根据不同的业务场景选择数据采集策略或工具； 4.1.2 能够使用工具采集工业网关数据 4.1.3 能够独立使用工具实现 MES 系统数据库与大数据系统之间的数据迁移； 4.1.4 能够独立实现 MES 系统日志数据的采集。
	4.2 工业数据清洗	4.2.1 能够根据业务需求，完成数据质量的评估； 4.2.2 能够根据业务需求，制定数据清洗的策略； 4.2.3 能够根据业务需求，运用数据清洗工具实现数据清洗。
5. 分布式数据库 HBase 管理与应用	5.1 分布式数据库部署与管理	5.1.1 能够根据工业大数据业务场景的需求，独立安装分布式数据库系统； 5.1.2 能够基于分布式数据库方案设计，实现分布式数据库的构建。
	5.2 分布式数据库的数据操作	5.2.1 能够独立对业务数据库进行数据查询、增加、删除操作。
6. 分布式数据仓库 Hive 管理与应用	6.1 分布式数据仓库部署与管理	6.1.1 能够根据工业大数据业务场景的需求，独立安装分布式数据仓库； 6.1.2 能够基于业务需求，实现分布式数据仓库的构建； 6.1.3 能够根据业务需求，选择合适的表类型存储数据 6.1.4 能够根据业务需求，实现表、视图对象的管理。
	6.2 分布式数据仓库的数据操作	6.2.1 能够根据业务需求，完成单表数据普通数据类型的条件查询； 6.2.2 能够根据业务需求，完成单表数据数组列、map、struct 特殊数据类型的条件查询； 6.2.3 能够根据业务需求，完成多表连接查询； 6.2.4 能够根据业务需求，使用数据函数、字符串函数、聚合函数等常用函数完成复杂的查询功能。 6.2.5 能够根据业务需求，独立对分布式数据仓库进行数据的 ETL 操作。
7. 内存数据库 Redis 管理与应用	7.1 内存数据库部署与管理	7.1.1 能够根据工业大数据业务场景的需求，独立安装内存数据库； 7.1.2 能够根据业务的需求，实现 Redis 主从复制的配置； 7.1.3 能够根据业务的需求，实现 Redis 持久化的设置。

	7.2 内存数据库的数据操作	7.2.1 能够使用内存数据库实时操作工业数据，进行数据的插入、查询、删除、选择数据库、监视实时转储收到的请求等操作； 7.2.2 能够对 strings、hash、lists、sets 等多种数据类型的数据进行处理。
8. 工业数据可视化	8.1 工业数据可视化基础	8.1.1 能够根据业务需求，针对不同的业务场景选择不同的图表展现形式； 8.1.2 能够根据数据可视化方案，选择适合的数据可视化工具。
	8.2 工业数据可视化实现	8.2.1 能够对选择的数据可视化工具进行基本的配置和操作 8.2.2 能够基于数据可视化方案设计，使用数据可视化工具进行数据的简单图表展示。

表 2 工业大数据应用与开发职业技能等级要求（中级）

项目名称	任务名称	任务描述
1. 工业大数据应用编程	1.1 Hadoop 数据开发代码编写	1.1.1 能够根据 Hadoop 开发要求，正确安装配置 Hadoop 开发所需要的环境； 1.1.2 能够通过团队协作，根据工业数据开发项目的需求，使用面向数据处理的开发的技术实现简单代码编写； 1.1.3 能够初步完成具有海量数据处理能力的代码编写。
	1.2 代码调试	1.2.1 能够根据程序语法规则，初步完成错误语法的处理； 1.2.2 能够初步进行代码单步调试、断点调试、输入输出调试与异常处理调试。
2. 分布式数据库 HBase 设计与开发	2.1 分布式数据库设计	2.1.1 能够根据业务需求，独立完成分布式数据库方案设计；
	2.2 分布式数据库的数据操作	2.2.1 能够根据业务需求，选择合适的过滤器进行数据的过滤操作； 2.2.2 能够根据业务需求，使用 Java 编程实现分布式数据库的数据操作； 2.2.3 能够根据业务需求，进行分布式数据库的批量处理。
3. 分布式数据仓库 Hive 编程与优化	3.1 分布式数据仓库文件存储与 UDF 编程实现	3.1.1 能够根据业务场景，设置不同的文件存储格式； 3.1.2 能够根据业务需求，完成用户自定义函数的应用；

		3.1.3 能够根据业务需求，完成 UDF 编程。
	3.2 分布式数据仓库性能优化	3.2.1 能够根据业务需求，选择合适的模式设计，实现数据的优化存储 3.2.2 能够根据数据的特征，进行 HQL 语句的优化。
4. 内存数据库 Redis 管理与应用	4.1 内存数据库的数据操作	4.1.1 能够实现 Java 到 Redis 服务器的连接； 4.1.2 能够通过连接池连接到 Redis 数据库； 4.1.3 能够使用 java 编程实现对多种数据类型的操作； 4.1.4 能够根据业务的需求，实现 Redis 集群的安装部署。
5. 工业数据可视化	5.1 工业数据可视化设计	5.1.1 能够根据业务需求，针对不同的业务场景选择不同的图表展现形式； 5.1.2 能够根据业务需求，完成工业数据可视化方案设计； 5.1.3 能够根据数据可视化方案，选择适合的数据可视化工具。
	5.2 工业数据可视化实现	5.2.1 能够对选择的数据可视化工具进行基本的配置和操作 5.2.2 能够基于数据可视化方案设计，使用数据可视化工具进行数据的复杂图表展示。
6. 时序数据库存储与应用	6.1 时序数据库部署与管理	6.1.1 能够根据工业大数据业务场景的需求，独立完成时序数据库安装部署； 6.1.2 能够实现时序数据库的权限管理； 6.1.3 能够实现时序数据库的备份、还原、导出、导入的管理。
	6.2 时序数据库的数据操作	6.2.1 能够根据业务需求，实现时序数据的增加、查询、删除操作； 6.2.2 能够根据业务需求，实现时序数据的连续查询； 6.2.3 能够根据业务需求，实现保留策略的创建、查询、管理等操作。
	6.3 时序数据库的应用实现	6.3.1 能够根据工业大数据业务场景完成 Telegraf 部署及设置； 6.3.2 能够根据工业大数据业务场景完成 Kapacitor 部署及设置； 6.3.3 能够根据工业大数据业务场景完成 Chronograf 部署及设置； 6.3.4 能够基于 TICK 技术栈搭建 DevOps 监控系统。

表 3 工业大数据应用与开发职业技能等级要求（高级）

项目名称	任务名称	任务描述
------	------	------

1. 工业大数据 Spark 内存批量处理	1.1 内存批量计算的软件代码编写	1.1.1 能够根据内存计算开发要求，正确安装配置 Spark 开发所需要的环境； 1.1.2 能够根据工业大数据功能模块需求，独立完成内存批量计算代码编写； 1.1.3 能够通过团队协作，根据工业大数据开发项目的需求，使用函数式编程技术实现各功能模块； 1.1.4 能够独立编写具有异常处理功能的内存批量计算代码模块。
	1.2 SQL 数据分析处理	1.2.1 能够根据业务需求，使用内存计算的 SQL 进行数据清洗； 1.2.2 能够根据业务需求，对多种数据源使用内存计算的 SQL 进行简单的数据分析； 1.2.3 能够根据业务需求，使用内存计算的 SQL 进行复杂的数据分析处理。
2. 工业大数据流式计算处理	2.1 消息队列 Kafka 开发	2.1.1 能够进行消息队列的安装与部署； 2.1.2 能够根据不同的业务需求进行单节点-单代理、单节点-多代理、多节点-多代理集群配置； 2.1.3 能够编写代码实现生产与消费。
	2.2 流式计算开发	2.2.1 能够构建流式计算的开发环境； 2.2.2 能够使用 TCP、HDFS、flume、消息队列等数据源进行流式计算的应用开发。
3. 工业大数据分析	3.1 数据清洗	3.1.1 能够通过编程的方式实现无效数据的清除； 3.1.2 能够通过编程的方式实现缺失数据的处理； 3.1.3 根据工业机理模型进行数据清洗； 3.1.4 根据业务的需要，对数据进行归一化处理。
	3.2 数据分析	3.2.1 能够对数据分析所使用的程序库进行安装； 3.2.2 能够使用扩展程序库进行维度数组与矩阵运算； 3.2.3 根据业务的需要，对数据进行合并、拼接、旋转、聚合处理。 3.2.4 能够使用数据分析包进行数据分析； 3.2.5 能够根据业务需求，从多种数据源读取数据进行数据分析； 3.2.6 能够根据业务需求，编写较为复杂的数据提取语句，准确地从多个数据表提取数据进行分析。
4. 工业大数据可视化	4.1 工业数据可视化设计	4.1.1 能够根据业务需求，针对不同的业务场景选择不同的图表展现形式； 4.1.2 能够根据业务需求，完成工业数据可视化方案设计； 4.1.3 能够根据数据可视化方案，选择适合的数据可视化工具。

	4.2 工业数据可视化实现	4.2.1 能够对选择的数据可视化工具进行基本的配置和操作 4.2.2 能够基于数据可视化方案设计,使用数据可视化编程技术进行数据的图表展示; 4.2.3 能够根据数据分析可视化结果,得出有效地分析结论。
5. 工业大数据的数据挖掘	5.1 数据挖掘	5.1.1 能够掌握数据挖掘整体实现过程; 5.1.2 能够根据业务场景,使用聚类分析模型的常用算法完成模型的构建; 5.1.3 能够根据业务场景,使用分类模型常用算法完成模型的构建; 5.1.4 能够根据业务场景,使用回归模型常用算法完成模型的构建; 5.1.5 能够根据业务场景,使用关联规则常用算法完成模型的构建。

参考文献

- [1] 中华人民共和国教育部,《中等职业学校专业目录》(2010年修订)[M].北京:高等教育出版社,2010-11.
- [2] 中华人民共和国教育部,《中等职业学校专业目录》增补专业(2019年)[Z].
- [3] 中华人民共和国教育部,《普通高等学校高等职业教育(专科)专业目录》(2015年)[Z].
- [4] 中华人民共和国教育部,《普通高等学校高等职业教育(专科)专业目录》2019年增补专业[Z].
- [5] 中华人民共和国教育部,《普通高等学校(本科)专业目录》(2012年)[Z].
- [6] GB/T 35295-2017 信息技术 大数据 术语
- [7] GB/T 35589-2017 信息技术 大数据 技术参考模型
- [8] GB/T 35274-2017 信息安全技术 大数据服务安全能力要求
- [9] GB/T 33745-2017 物联网 术语
- [10] GB/T 5271.1-2000 信息技术 词汇 第1部分:基本术语
- [11] GB/T 38619-2020 工业物联网 数据采集结构化描述规范
- [12] GB/T 32919-2016 信息安全技术 工业控制系统安全控制应用指南
- [13] SJ/T 11666.1—2016 制造执行系统(MES)规范 第1部分:模型和术语