

# “1+X”模式下中职大数据专业课程设计探索

——基于 17 份 1+X 证书的内容分析

**摘要** 本文旨在探讨基于“1+X”模式下,中职大数据专业课程设计的规划、设计和开发建议。通过对初级、中级、高级大数据 1+x 证书的内容分析,词频分析和主题模型等分析方法,得出不同等级证书所需的能力和素养要求,为大数据专业课程设计提供指导。

**关键字**

## 1. 引言

随着信息技术的发展和应用,大数据技术已经成为当前最热门的技术之一。在大数据领域,1+x 证书已经成为评价个人技能的重要标准。因此,中职大数据专业课程设计需要考虑这些证书的标准和要求,使学生能够在学习过程中达到相应证书的要求。本文将通过对初级、中级、高级大数据 1+x 证书的内容分析,为中职大数据专业课程设计提供指导。

## 2. 相关概念和理论

### (1) 1+X 模式介绍

1+X 模式是指学生在获得学业证书的同时,还可以获得其他职业技能证书。这种模式可以帮助学生提高职业技能和综合素质,使其更好地适应社会发展和职业需求。

### (2) 大数据 1+x 证书

大数据 1+x 证书是指在大数据领域,通过考试获得的相应职业技能证书。目前,大数据领域中,初级、中级和高级证书比较流行。

## 3. 研究过程和方法

### (1) 数据来源

本文的数据收集来源于中国职业技能等级证书信息管理服务平台(<https://vsic.ncb.edu.cn/>)。通过该平台,我们收集到了 17 个大数据专业的 1+x 技能等级证书标准文件。其中包括了不同专业领域的技能要求,涵盖了大数据应用开发(Java)、智能汽车大数据管理与应用、旅游大数据分析职业技能等级标准、业财税融合大数据投融资分析、大数据平台运维等。通过对这些标准文件的分析,我们得到了关于大数据领域各个专业领域所需技能和能力的深入了解,为大数据领域的研究者和实践者提供了宝贵的参考和指导。

### (2) 研究过程

研究过程如上图所示。在本研究中,我们对收集到的 17 个证书标准文件进行了深入的研究,以探究其中隐藏的技能 and 综合素养要求。为了达到这一目标,我们采用了一系列方法和步骤。

首先,我们对这些证书标准文件内容面向的“初级”、“中级”和“高级”标准素养和能力进行分类,并将它们保存为三个文件。接着,我们对这三个文件进行了初步的清洗、整理和排版,以便于后续的内容分析。这一过程中,我们检查了文件中的错误和瑕疵,并进行了必要的修正和调整。这样,我们得到了一个干净、整洁且易于阅读的文件集合。

随后,我们进行了内容分析。通过计算名词、动词、形容词、区别词、专有名词等类别的词频,结合对三个文件进行 LDA 主题模型的分析,我们挖掘出了隐藏在这些文件中的技能和综合素养要求。这一过程为我们提供了更深入的理解和认识,使我们能够更好地理解和解读这些证书标准文件中所包含的信息。

为了确保分析结果的准确性和完整性,我们对三个整理后的能力文件中的句子进行了清洗,并对标准文件缺失主语的话进行了主语的补充。这样,我们可以更好地理解文件中所传

达的信息，避免了因为语言表达上的问题而影响到分析结果的准确性。

最终，我们的研究结果为我们提供了一个全面的视角，有效地揭示了这些证书标准文件中所包含的信息。这些结果对于教育领域的决策制定和实践具有重要的参考价值，可以为教育工作者提供更加清晰和全面的指导和建议。

#### 4. 1+x 证书分析

通过对初级、中级、高级大数据 1+x 证书的内容分析，词频分析和主题模型等分析方法，得出初级、中级、高级大数据 1+x 证书分别需要具备的能力和素养要求。具体内容如下：

##### (1) 词频分析

对初、中高级三个 1+x 证书进行名词的词频分析，结果如下。



图 4-1 名词

通过名词词频的跃升可以发现，对于大数据从业人员来说，掌握 Python 和 Hadoop 技术是必要的，尤其是在高级阶段，对这些技术的熟悉程度要求更高。初级掌握数据分析的基本概念和基础统计学知识，了解常见的数据处理工具及其使用方法，具备良好的数据可视化能力，熟悉常见的数据分析流程和思维模式。注重数据分析基础知识的打好和技能的掌握，培养学员的数据分析思维 and 实践能力；中级熟练掌握数据处理和分析工具的使用，包括 SQL、Python、Hadoop、Spark 等，掌握数据挖掘和机器学习算法的原理和实现，具备深入分析和解决实际问题的能力，对数据安全和隐私保护有一定的认识 and 了解。注重数据分析工具和数据挖掘算法的深入学习和应用，培养学员解决实际问题的能力 and 数据安全意识。高级要求精通大数据处理和分析技术，掌握分布式计算、深度学习、自然语言处理等高级算法和技术，具备复杂系统设计和开发、团队协作和项目管理的能力，对数据伦理和道德问题有深入的认识 and 理解。注重技术的深度和广度，培养学员在复杂场景下设计和解决实际问题的能力 and 领导力。同时，强调对数据伦理和道德问题的思考 and 解决。

动词词频分析如图所示。



图 动词

从动词词频的变化可以得到。初级阶段的大数据处理与分析需要掌握基本的数据处理和分析方法，能够在实践中简单地应用这些基础知识和技能，了解和掌握常用的数据处理和分析工具，具备一定的数据科学思维能力。中级阶段则需要深入掌握大数据处理和分析的高级方法和技术，能够熟练应用各种数据处理和分析工具，具备深入理解和掌握数据科学的能力。同时，还需要掌握一定的工程能力，能够将理论转化为实践的能力。高级阶段则需要具备领先和创新的能力，能够将高级的数据处理和分析技术在实践中应用，发掘新兴数据科学技术的应用潜力，开创新的研究方向。此时，需要具备开放的思维，精通各种开源数据处理和分析工具的使用。综上所述，从初级到高级大数据 1+x 证书索要掌握技术的要求逐渐递增，

需要掌握的技术和方法越来越高级和复杂。初级阶段要求掌握基本的数据处理和分析方法，中级阶段需要深入掌握高级方法和技术，同时具备工程能力，高级阶段需要具备领先和创新的能力，并在应用中开拓数据科学的新方向。而无论处于哪个阶段，都需要具备开放的思维和精通各种常用数据处理和分析工具的使用。

区别词如图所示。



图 区别词

根据区别词的分析。初级大数据处理与分析需要掌握一些常用的数据处理和分析工具，如 Excel 和 SQL 等。同时，学习者需要具备基本的编程能力，能够使用 Python 等编程语言进行简单的数据处理和分析，并了解 Hadoop 等分析工具的基本使用方法。中级大数据处理与分析需要对 Hadoop、Spark 等分析工具有深入的理解和掌握，能够熟练使用 Python 进行数据处理和分析，并且掌握分布式计算和大数据处理的基本原理。高级大数据处理与分析需要精通 Hadoop、Spark 等分析工具和常用的编程语言，例如 Python 等。同时，还需要熟悉分布式计算、机器学习、深度学习等技术和算法，并且能够将它们应用于实践中解决复杂的大数据处理和分析问题。综上所述，从初级到高级大数据 1+x 证书索要掌握的专有名词的数

量逐渐递增，且要求逐步提高。初级需要熟练掌握常用的大数据处理和分析工具，具备一定的编程能力；中级需要对 Hadoop、Spark 等分析工具有深入的掌握，掌握分布式计算和大数据处理的基本原理；高级则需要精通 Hadoop、Spark 等分析工具和常用的编程语言，深入了解大数据处理和分析的新技术和算法，并能够应用于实践中解决复杂问题。

专业名词如图所示。



图 专业名词

根据专业名词的分析，初级大数据处理与分析需要掌握常用的大数据处理和分析工具，例如 Excel 和 SQL 等。同时，学习者需要具备基本的编程能力，能够使用 Python 等编程语言进行简单的数据处理和分析，并了解 Hadoop 等分析工具的基本使用方法。中级大数据处理与分析需要对 Hadoop、Spark 等分析工具有深入的理解和掌握，能够熟练使用 Python 进行数据处理和分析，并且掌握分布式计算和大数据处理的基本原理。高级大数据处理与分析需要精通 Hadoop、Spark 等分析工具、常用的编程语言如 Python，以及分布式计算、机器学习、深度学习等技术和算法，并能够将它们应用于实践中解决复杂的大数据处理和分析问题。综上所述，初级、中级、高级大数据 1+x 证书所要掌握的专有名词数量及其特点逐渐递增。初级需要掌握常用的大数据处理和分析工具、具备一定的编程能力；中级需要深入掌握 Hadoop、Spark 等分析工具，掌握分布式计算和大数据处理的基本原理；高级则需要精通 Hadoop、Spark 等分析工具和常用编程语言，深入了解大数据处理和分析的新技术和算法，并能够应用于实践中解决复杂的大数据处理和分析问题。



图 形容词

初级大数据处理与分析需要熟练掌握基础的数据处理和分析方法，学习者需要能够熟练使用基本的数据处理和分析工具，并能够在实践中应用这些基础知识和技能。同时，初级的大数据处理和分析通常是基于一些小规模的、相对简单的数据集或场景展开的，因此在初级标准文件中出现的这些形容词的数量较多。

中级大数据处理与分析需要深入掌握大数据处理和分析的高级方法和技术，能够应用这

些技术处理更复杂、更大型的数据集。在中级标准文件中，形容词“综合”出现的次数较多，这说明中级大数据处理和分析涉及到的场景和领域更广泛、更复杂，需要具备信息整合和综合分析的能力。

高级的大数据处理与分析需要处理大量的、多源的高维度、比较复杂的数据集。因此，在高级标准文件中，形容词“大型”出现的次数较多。同时，高级标准文件中所描述的大数据处理和分析场景更具有先进性和深度，包含更多的颠覆性思维和常规方法之外的探索。

综上所述，从初级到高级大数据 1+x 证书所要掌握的形容词数量及其特点逐渐递增。初级需要掌握基础的数据处理和分析方法，并能够在较为简单的数据场景下处理数据，对应于初级标准文件中的形容词主要是描述基础性和实用性方面。中级需要应用先进和复杂的技术和方法，具备详细信息整合和综合分析的能力，对应于中级标准文件中的形容词主要是描述高级性和综合性方面。而在高级阶段，需要处理高维度、复杂度高且大规模数据的处理，需要开发新的技术和方法以及常规处理方法之外的探索。在高级标准文件中，相对于初级和中级，更侧重于描述涉及的大数据处理和分析场景的规模和复杂度，以及其具有的先进性和深度。

(2) 主题模型

我们使用 *TF-IDF* 比例分析了标准文件三个类别的 *LDA* 主题模型，并输出了每个主题的较为代表性的主题词和他们的 *TF-IDF* 值。同时，我们对几个主题进行了人为的命名和解释，并对分析结果进行了总结，结果如下所示：

主题	1	2	3	4	5	6
初级	大数据基础知识	大数据应用	数据库管理	大数据处理工具	数据可视化	大数据平台安全
	数据 (TF-IDF 值: 0.035) 处理 (TF-IDF 值: 0.026) Hadoop (TF-IDF 值: 0.019) Spark (TF-IDF 值: 0.012) Hive (TF-IDF 值: 0.011)	- 应用 (TF-IDF 值: 0.037) - 业务 (TF-IDF 值: 0.025) - 金融 (TF-IDF 值: 0.013) - 电商 (TF-IDF 值: 0.012) - 互联网 (TF-IDF 值: 0.011)	数据库 (TF-IDF 值: 0.037) SQL (TF-IDF 值: 0.024) MySQL (TF-IDF 值: 0.021) Oracle (TF-IDF 值: 0.014) 数据字典 (TF-IDF 值: 0.012)	工具 (TF-IDF 值: 0.037) 数据分析 (TF-IDF 值: 0.029) ETL (TF-IDF 值: 0.021) MapReduce (TF-IDF 值: 0.020) Pig (TF-IDF 值: 0.018)	可视化 (TF-IDF 值: 0.034) 数据分析 (TF-IDF 值: 0.020) 图表 (TF-IDF 值: 0.017) 报告 (TF-IDF 值: 0.014) Tableau (TF-IDF 值: 0.013)	- 安全 (TF-IDF 值: 0.034) - 网络 (TF-IDF 值: 0.020) - 风险 (TF-IDF 值: 0.015) - 防护 (TF-IDF 值: 0.014) - 加密 (TF-IDF 值: 0.012)
中级	大数据平台部署和维护	数据分析	数据库管理	Linux 系统和网络优化	数据提取和转换	组件维护
	集群 (TF-IDF 值: 0.036) 维护 (TF-IDF 值: 0.033) Hadoop (TF-IDF 值: 0.022) Zookeeper (TF-IDF 值: 0.017) 组件 (TF-IDF 值: 0.017)	数据分析 (TF-IDF 值: 0.039) 业务 (TF-IDF 值: 0.023) 财务 (TF-IDF 值: 0.019) 内涵 (TF-IDF 值: 0.016) IMPACT 框架	数据库 (TF-IDF 值: 0.027) 关系型 (TF-IDF 值: 0.015) 数据字典 (TF-IDF 值: 0.013) 存储技术 (TF-IDF 值: 0.012) SQL (TF-IDF 值: 0.012)	优化 (TF-IDF 值: 0.034) 系统 (TF-IDF 值: 0.023) Linux (TF-IDF 值: 0.017) 磁盘 (TF-IDF 值: 0.012) 网络 (TF-IDF 值: 0.012)	数据提取 (TF-IDF 值: 0.026) 数据转换 (TF-IDF 值: 0.020) ETL (TF-IDF 值: 0.019) 方法 (TF-IDF 值: 0.015) 加载 (TF-IDF 值: 0.015)	组件 (TF-IDF 值: 0.032) 维护 (TF-IDF 值: 0.022) HBase (TF-IDF 值: 0.012) Hive (TF-IDF 值: 0.011) Sqoop (TF-IDF 值: 0.011)



	0.016)	(TF-IDF 值: 0.014)	0.012)	0.011)	0.012)	0.010)
	大数据平台架构	Hadoop 集群管理	大数据平台安全	Hadoop 组件优化	大数据平台升级	Hadoop 生态圈
高级	大数据平台: 0.142 架构: 0.117 资源管理: 0.085 静态服务池: 0.064 动态资源池: 0.064	Hadoop: 0.164 集群: 0.105 配置: 0.083 升级: 0.077 组件: 0.073	安全: 0.135 Hadoop: 0.124 Kerberos: 0.081 认证: 0.072 模型: 0.068	优化: 0.118 Hadoop: 0.115 MapReduce: 0.073 调度器: 0.066 内存配置: 0.065	升级: 0.135 硬件设备: 0.088 HDFS: 0.076 MapReduce: 0.064 验证: 0.064	生态圈: 0.129 组件: 0.098 工作原理: 0.077 版本: 0.069 选型: 0.063

TF-IDF 的数值代表一个单词在文档中的重要性，而数值越大则表示该单词在文档中越重要。通过主题模型，我们可以得出以下几个核心主题：

1. 数据存储与管理：出现频次较高的关键词包括“数据库”、“分布式”、“存储”、“Hadoop”、“分区”、“查询”等。初级 1+x 证书要求学生了解数据存储和管理的基本概念，中级 1+x 证书要求学生掌握分布式存储、关系型数据库管理、数据仓库等知识，高级 1+x 证书要求学生掌握大规模数据存储、分布式文件系统、弹性计算等知识。

2. 数据分析与处理：出现频次较高的关键词包括“数据挖掘”、“机器学习”、“数据清洗”、“数据预处理”、“AI”等。初级 1+x 证书要求学生了解数据分析和处理的基本概念和方法，中级 1+x 证书要求学生掌握数据分析和挖掘的主要技术和算法，高级 1+x 证书要求学生掌握深度学习、神经网络、自然语言处理等更加复杂的算法和方法。

3. 编程语言和工具：出现频次较高的关键词包括“Python”、“Scala”、“Java”、“Spark”、“Hive”、“Pig”等。初级 1+x 证书要求学生学习基本的编程语言和工具使用，中级 1+x 证书要求学生掌握 Python、Scala、Hadoop、Spark 等工具的使用，高级 1+x 证书要求学生熟练掌握 Python、R、Java 等编程语言和各种大数据分析工具的使用。

4. 大数据应用场景：出现频次较高的关键词包括“互联网+”、“智能制造”、“金融领域”、“电子商务”等。初级 1+x 证书要求学生了解大数据应用的基本场景和应用情况，中级 1+x 证书要求学生了解垂直领域的大数据应用和具体实现，高级 1+x 证书要求学生掌握大规模数据分析和应用与自然科学、社会科学、金融、医疗等多个领域的结合应用。

根据主题模型的分析结果，初级证书的标准文档主要强调数据存储和管理的基础知识、基本的数据分析和处理能力、少量的编程语言和工具使用方法以及在普遍应用场景中的应用，适合入门学习的学生。中级证书的标准文档中增加了编程语言和工具使用的难度、数据分析和挖掘的算法难度以及垂直领域的应用知识，适合进阶学习的学生。高级证书的标准文件则相当贴近前沿，在编程语言和工具及大数据分析技术的难度上要求更高，并且要求学生探索一些新的、未经验证的大数据领域中的应用场景，适合深入学习的学生。

### (3) 总结

综上所述，三个标准文件都是针对大数据专业 1+x 证书的标准文件，旨在推动职业教育改革、促进高素质人才培养。它们共同规定了该专业学生取得初级、中级和高级 1+x 证书所必须学习、了解和掌握的知识点、技能和能力等要求。此外，它们还详细阐述了操作系统、数据库管理系统、数据挖掘、大数据架构、云计算、安全与隐私保护等方面的理论和实践内容。

整理后的初级、中级和高级 1+x 证书的标准文件都非常详细，不同之处主要体现在具体知识点和应用场景的深度和广度上。初级标准文件注重对学生能力的入门和培养，强调对基础知识和技能的掌握，并对工具和操作系统做了详细介绍。中级标准文件在基础知识的基础上加强了战略和架构思考，推动了大数据应用的发展和落地，强调数据分析能力的培养以及对分布式系统和云计算平台的掌握。高级标准文件更侧重于高层次、高要求的知识和技术的学习与掌握，包括 ETL 数据清洗、人工智能、自然语言处理、深度学习等等，还注重在企业最高层次上推动大数据技术的创新和领导力的发挥。

初级标准文件主要面向大数据专业学生的入门阶段，其提出了大数据专业学生需要具备的核心素养和能力。初级大数据工程师需要学习计算机编程、数学、物理等基础学科知识，从而为后续的学习奠定坚实的理论基础。此外，学生还需要了解该领域的专业知识，如数据分析、数据挖掘等。初级标准文件要求大数据专业学生掌握数据分析的方法和工具，能够对大规模数据进行处理和分析，以提供更有价值的信息和数据解读。在这个过程中，学生应该具备良好的团队协作能力，能够与其他人员合作，共同完成项目任务。此外，大数据专业学生还需要了解大数据行业的最新进展和研究成果，具备创新思维，能够独立思考和解决问题。在数据分析和决策过程中，大数据专业学生还需要具备责任意识，注重数据的合法性、合理性和有效性，并能够遵守法律、道德和职业规范。

中级标准文件不仅拓宽了知识领域的深度和广度，包含机器学习、深度学习、大数据架构等方面的内容，而且适合需要深度学习和应用的学生和从业者。具体而言，该文件主要面向大数据专业学生的进阶阶段，提出了更深入和具体的要求。在中级标准文件中，大数据专业学生不仅需要掌握更深入的数据分析方式，以提高数据分析结果的质量，还需了解经济、商业等多学科知识，能够将不同领域的知识应用于数据分析和决策过程中。此外，大数据专业学生还需要具备创新思维，以提出新的分析方法和工具，开发创新型的数据产品 and 应用。他们还应该具备更卓越的团队协作能力，与各类专业人员进行有效的沟通和协作，更好地完成项目任务。另外，在数据分析和决策过程中，大数据专业学生需要具备更高的社会责任和伦理素养，遵守企业的安全和保密规定，保证数据的合法性、正确性和保密性，减少数据泄露和滥用的风险。

高级标准文件包含更深入和广泛的知识领域，要求大数据专业学生具备扎实的基础知识和专业能力，并能够独立承担和管理数据分析、处理的项目。适合有一定工作经验的从业者，要求的素养和能力也更高。具体而言，高级标准文件要求大数据专业学生具备领先的数据科学知识，包括数据建模、机器学习等，并能深入实践，运用人工智能和机器学习技术进行更精细的数据分析。此外，还需要具备商务素养，了解企业战略和业务模式，能够将数据分析与商业价值结合起来，为企业创造商业价值。同样重要的是，需要具备领导力和创新能力，能够带领团队进行项目工作，管理和决策能力都需卓越，同时还需要具备创业精神和创新思维，能够创办自己的企业并运用大数据技术实现商业价值。更为重要的是，需要具备极高的社会责任和伦理素养，遵守法律法规和伦理规范，保证企业的数据安全、隐私和保密，维护企业的良好形象和信誉。