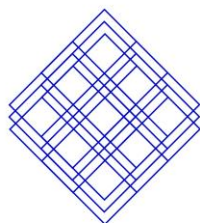




Enterprise Big Data Professional®



ENTERPRISE BIG DATA PROFESSIONAL

教学大纲

2018.04

[本中文大纲由上海北宙企业管理咨询有限公司刘颀翻译]

介绍

大数据框架 (Big Data Framework) 为企业提供了一体化的精编方法，旨在帮助能更好的挖掘组织中的数据价值。本框架涵盖了理解和分析海量数据所必需的大数据的方方面面的基本要素。

对应这个大数据框架的企业大数据专家级认证 (Enterprise Big Data Professional qualification)，可令考生对大数据有基本的了解，并且能够理解大数据框架的结构和核心元素。企业大数据专家级认证是大数据框架认证 (Big Data Framework qualification) 体系的第一个级别，涵盖了与大数据相关的基本知识和基本概念。

本教学大纲涵盖了企业大数据专家级认证考试 (Enterprise Big Data Professional examination)。它基于大数据框架网站 (www.bigdataframework.org) 上发布的第一版企业大数据专家指南 (“文本”)，适用于参加考试的所有考生。

教学大纲的主要目的是为参与大数据解决方案和服务的人员提供认证的基础，它描述了与使用大数据框架相关的学习成果，并描述了各个认证级别所期望考生达到的学习成果的范围。

这份文件适用于：

- 考试委员会
- 考试小组
- APMG 评估小组
- 授权培训机构
- 企业大数据专家级考试考生

这一教学大纲说明了有关考试设计的信息，并为经授权的培训组织提供了考试评估内容的更详细的分解信息。关于考试结构和内容的细节在*企业大数据专家级考试设计*中进行记录。

1. 企业大数据专家级认证 (Enterprise Big Data Professional Qualification)

1.1 目的

企业大数据专家级认证目的是衡量考生是否对大数据框架有足够的知识和理解，以及考生是否能够理解基本数据处理技术和算法，并以此解决实际问题。

经过认证的企业大数据专家能分析实践和并掌握技术概念，这些概念包括当前大数据环境和工具中的特征。

1.2 目标群体

这个认证针对参与企业大数据分析的人员，他们需要了解大数据背后原理的工作知识，并需要知道相关术语和实践背后的一些理论。因此，企业大数据专家级认证的目标受众包括：

- 数据分析师 (Data Analysts)
- 业务分析师 (Business Analysts)
- IT 顾问 (IT Advisors)
- IT 专家 (IT Professionals)

企业大数据专家级认证是所有希望了解大数据术语、核心概念和大数据解决方案的基本知识专家的入门级课程。

该课程将提供现在主要技术的概览，但不会涉及编程或实施，获得企业大数据专家级认证没有强制性的先决条件。

1.3 高水平的成功考生定义

满足高水平定义的考生应该至少能够记忆并识别和展示在《企业大数据专家框架指南 (Enterprise Big Data Professional Framework text) 》中阐述的理论、设计元素、过程和技术的理解。此外，一个成功的考生应该能够理解基本的数据处理技术和算法，以解决实际问题。

具体地说，他（她）应该能够表明已经理解了一下内容：

- 1) 解释大数据的历史和背景及其与传统的数据分析技术和解决方案的差异；
- 2) 描述数据分析、大数据、机器学习和人工智能的基本特征以及这些概念之间的差异。
- 3) 描述如何制定大数据战略，如何建立支撑大数据和大数据解决方案的业务驱动计划，以获得大数据的价值。
- 4) 讨论当代大数据架构的高级原则和设计元素，并解释它们的核心收益。
- 5) 解释基本的大数据算法和处理技术，以便选择适当的技术来解决实际的大数据问题。
- 6) 识别如何应用大数据流程并根据企业组织的需要去从海量数据中去捕获价值。
- 7) 解释组织在从大数据中获取长期价值所必需的关键职能、角色和权限。
- 8) 了解人工智能的重要性和概念及其与大数据方法和大数据解决方案的关系。

2. 评估模型

高水平的学习成果是要求考生能去展现出其具备特定的知识和技能。对于每个学习成果，我们确定了一些学习结果的测量方法，通过考试设计在考试中进行评估，以此来确认学习成果已经实现。这些学习成果措施会显示为教学大纲课题，以此界定达到资格所需的标准范围。

在设计认证和学习的评估时广泛使用的方法是布鲁姆教育目标分类法 (Bloom' s Taxonomy of Educational Objectives)。这将学习目标划分为六个学习级别，每个层次都定义了更高的能力和技能。(布鲁姆，1956，教育目标分类学)。

APMG 已将其纳入学习成果的评估模型，然后用于开发每个资格的评估模型。该模型为评估和分类学习结果措施提供了一种简单而系统的方法。

这种结构化方法有助于确保：

- 为认证定义合适的级别
- 清晰界定不同认证的学习水平和内容
- 措辞标准化，确保教学大纲在 APMG 的认证组合中出现是一致的
- 考题和考卷在设计上是一致的。

目前认证审主要查基于水平 1 (记忆) 和 2 (了解)。

大数据框架评估模型				
	1. 记忆(Recall)	2.了解(Understand)	3.应用(Apply)	4.分析(Analyze)
APMG 学习水平定义	<i>记住以前学过的信息</i>	<i>掌握信息的含义和意义</i>	<i>使用信息来执行技能或任务</i>	<i>根据规则和指导，确定是否正确使用了信息</i>
通用 APMG 课题 <i>介绍教学大纲中学习结果的考量 (主题)</i>	根据参考材料记忆有关概念、原则和过程的术语和关键要素	根据参考材料理解关键的要素、概念、原则和过程	将关键要素、概念、原则和过程应用于给定的场景	在给定的场景中区分参考材料的适当和不适当的使用
认证示例	记忆与教学大纲领域相关的概念、技术、设计元素和过程的术语和关键要素	理解与教学大纲领域相关的概念、技术、设计元素和过程	将与教学大纲区域相关的特定概念、技术、设计元素和过程应用于给定的场景	将与教学大纲区域相关的特定概念、技术、设计元素和过程的适当和不适当使用区分为给定的场景

3. 认证范围

每个认证的范围定义会在本文件末尾的教学大纲表格中列出，每个教学大纲区域都是与参考材料或培训课程模块相关的学习单元。

确定了下列教学大纲领域。

代码	标题
CO	Big Data Concepts and Key Drivers 大数据概念和关键驱动因素
FR	The Big Data Framework 大数据框架
ST	Big Data Strategy 大数据战略
AR	Big Data Architecture 大数据架构
AL	Big Data Algorithms 大数据算法
PR	Big Data Processes 大数据流程
FU	Big Data Functions 大数据职能
AI	Artificial Intelligence 人工智能

教学大纲介绍

每个教学大纲的区域中学习成果的考量需要按照学习层级的顺序呈现，所以会采用标准化的标题，每个教学区域的学习层级只有一个标题，该标题中的用语来源于评估模型，每次考量都会针对特定的学习层级进行。

在课题描述的右边都会有记号显示每个考试范围。

从业者资格认证包括了基础级别和从业者资格级别学习成果的考量的总和。从业者资格认证要求首先具备基础级别认证，在从业者资格认证中假定已满足这个条件不会再直接评估，但是基础级别的知识理解还是会被用于反映从业者资格应用和分析学习成果中。

每个教学大纲区域都以类似的格式呈现，如下所示：

区域代码 CO [2]	教学大纲区域： BD 基本概念 (CO) [1] 关键概念	主要参考章节
层级	课题	
熟记与大数据相关的关键术语和定义：[3] 需要记忆		

区域代码 CO [2]		教学大纲区域： BD 基本概念 (CO) [1]关键概念	主要参考章节	
01 [4]	01 [5]	[6] 大数据的定义		[7] 1.1
01	02			

教学大纲区域表的关键字

- | | | |
|---|--------------|-------------------------------|
| 1 | 教学大纲区域 | 学习单元，例如：课程模块、主要活动区域或参考指南的章节。 |
| 2 | 教学大纲区域代码 | 用于识别教学大纲区域的唯一的 2 字符代码。 |
| 3 | 学习层级标题 | 标题为给定的学习层级引入教学大纲主题(学习成果的考量域)。 |
| 4 | 层级 | 学习成果考量的学习层级。 |
| 5 | 课题参考 | 学习层级内的课题编号。 |
| 6 | 课题描述(考量学习成果) | 准确和具体地描述考生需要通过什么来证明学习成果已经达到了。 |
| 7 | 主要参考章节 | 支持学习成果考量的主要参考章节。 |

4. 要点

关于教学大纲的使用，应注意以下几点。

4.1 企业大数据专家指南参考 (Enterprise Big Data Professional Guide Reference)

所提供的企业大数据专家指南参考文献应被认为是指示性的，而不是全面的，也就是说在指南中可能会有其他有效的参考资料。

区域代码 CO		教学大纲区域: 大数据关键概念	主要参考章节
层级	课题		
记忆与大数据有关的关键术语和定义 需要记忆：			
01	01	大数据的定义	1.1
01	02	大数据的四个特征的名称	1.4
01	03	两类机器学习的名称和通常与之相关的技术: 1.监督式 - 分类和回归 2.非监督式 - 聚类和关联	1.7
了解大数据的起源及其关键概念的特点 需要理解：			
02	01	大数据的起源和三类大数据发展阶段的特点： 1.阶段 1 2.阶段 2 3.阶段 3	1.3
02	02	大数据的四个特征以及它们如何区分大数据和传统数据分析： 1. Volume – 体量 2. Velocity – 速率 3. Variety- 多样性 4. Veracity – 可信度	1.4
02	03	四种识别模式： 1. analysis - 分析 2. analytics- 解析 3. business intelligence - 商业智能 4. Big Data - 大数据	1.5
02	04	不同类型的解析目的： 1. descriptive - 描述性 2. diagnostic - 诊断性 3. predictive - 预测性 4. prescriptive - 规范性	1.5
02	05	大数据环境中元数据的功能	1.6

区域代码 CO		教学大纲区域: 大数据关键概念	主要参 考章节
02	06	三类数据类型的特征： 1. Structured - 结构化 2. Unstructured - 非结构化 3. Semi-structured - 半结构化	1.6
02	07	Hadoop 在分布式存储和分布式处理中的作用	1.7
02	08	了解两类机器学习，并能够识别相关样例： 1. Supervised - 监督式 2. Unsupervised - 非监督式	1.8

区域代码 FR		教学大纲区域: 大数据框架	主要参 考章节
层级	课题		
		记忆大数据框架相关的术语和关键要素 需要记忆：	
01	01	大数据框架六种能力的名称	2.2
		了解大数据框架的结构 需要理解：	
02	01	如何建立大数据组织的六大数据框架能力的相关性	2.2
02	02	大数据成熟度模型的不同层次: 1. Level 1 - Analytically Impaired – 分析能力有损级 2. Level 2 - Localized Analytics – 分析能力本地级 3. Level 3 - Analytical Operation - 分析能力运营级 4. Level 4 - Analytical Enterprise - 分析能力企业级 5. Level 5 - Data Driven Enterprise - 数据驱动企业级	2.4

区域代码 ST		教学大纲范围: <i>大数据战略</i>	主要参考章节
层级	课题		
记忆有关的大数据战略的关键要素 需要记忆			
01	01	制定大数据战略的五个步骤及其顺序	3.1
了解如何制定大数据战略以及所涉及的活动和技术 需要理解			
02	01	制定大数据战略所需要的六个业务驱动影响因素，以及如何使用大数据来产生竞争优势	3.1 and 3.2
02	02	优先级矩阵 1. 目的 2. 结构	3.3
02	03	制定大数据战略五个步骤中的每一步所涉及的活动: 1. 步骤 1-定义业务目标 2. 步骤 2-评估当前现状 3. 步骤 3-识别用例并排优先级 4. 步骤 4-制定大数据路线图 5. 步骤 5-通过变更管理植入	3.3

区域代码 AR		教学大纲区域: <i>大数据架构</i>	主要参考章节
层级	课题		
记忆大数据架构相关的术语和关键要素 需要记忆 :			
01	01	指导性架构与其目标	4.2

区域代码 AR		教学大纲区域: 大数据架构	主要参考章节
01	02	NIST 大数据参考架构的主要特点: 1. 整体结构 (5 个逻辑角色和 2 个维度) 2. 角色名称 3. 维度名称 4. 信息如何在不同角色之间流动	4.2
01	03	Hadoop 架构下核心组件的名称: 1. NameNode – 名字节点 2. MapReduce – 映射归约 3. SlaveNode - 从属节点 4. Job tracker – 工作跟踪器 5. HDFS - 分布式文件系统	4.6
了解当前大数据架构的高级原则和设计元素 需要理解:			
02	01	使用大数据参考架构的收益	4.2
02	02	与参考架构中的逻辑角色相关联的职能和活动 1. System Orchestrator - 系统编配器 2. Data Provider - 数据提供者 3. Big Data Application Provider - 大数据应用提供者 4. Big Data Framework Provider - 大数据框架提供者 5. Data Consumer - 数据消费者	4.2
02	03	本地式处理和分布式存储处理的区别	4.3
02	04	面对海量数据的三种大数据存储系统: 1. Direct Attached Storage (DAS)- 直接附加存储 2. Network Attached Storage (NAS) - 网络附加存储 3. Storage Area Network (SAN) - 存储区域网络	4.4
02	05	大数据存储机制 1. File systems - 文件系统 2. NoSQL databases - NoSQL 数据库 3. Parallel programming models - 并行编程模型	4.4

区域代码 AR		教学大纲区域: 大数据架构	主要参考章节	
02	06	大数据实证分析体系结构: 1. Real time analysis- 实时分析 2. Off-line analysis - 离线分析		4.5
02	07	Hadoop 在大数据环境中的功能		4.6
02	08	以下 Hadoop 组件的角色 1. NameNode – 名字节点 2. MapReduce – 映射归约 3. SlaveNode - 从属节点 4. Job tracker – 工作跟踪器 5. HDFS – 分布式文件系统		4.6

区域代码 AL		教学大纲区域: 大数据算法	主要参考章节	
层级	课题			
记忆大数据算法和分析技术的相关术语和关键要素				
需要记忆：				
01	01	什么是描述性统计信息		5.2
01	02	关联的关键要素 1. 什么是关联 (correlation) 2. 两种用于关联的变量类型 3. 关于 Pearson 皮尔逊关联系数的关键要素： - 它用来测量什么 - 其取值范围 - 负值，正值或 0 的含义	5.4	
01	03	分类的关键要素 1. 它能做什么 2. 机器学习的形式是什么？	5.6	

区域代码 AL		教学大纲区域: 大数据算法	主要参考章节
层级	课题		
了解大数据的基本算法和分析技术			
需要理解：			
02	01	对于每种类型的描述性统计，了解每个统计操作/分布措施或显示: 1. Central tendency statistics - 集中趋势统计 2. Dispersion statistics and – 离散统计 3. Distribution Shapes – 分布形态	5.2
02	02	偏度特征 1. Positive - 正偏度 2. Negative – 负偏度	5.2
02	03	大数据计算中为什么要标准化	5.2
02	04	识别和计算描述性统计的示例	5.2
02	05	不同类型的分布式图形的不同类型特征: 1. Frequency - 频率分布 2. Probability - 概率分布 3. Sampling – 抽样分布 4. Normal – 正态分布	5.2
02	06	为什么分布式图形对大数据和数据科学很重要: 1. Probability - 概率分布 2. Sampling – 抽样分布 3. Normal – 正态分布 4. Skew – 偏度	5.2
02	07	人口、抽样和偏倚对大数据的影响	5.3
02	08	如何在大数据中使用关联，并识别这些示例。	5.4
02	09	关联与回归的区别	5.5
02	10	识别分类算法的示例	5.6
02	11	聚类的关键特性: 1. 它能做什么 2. 大多数聚类算法的典型着眼点	5.7
02	12	如何在大数据的上下文中使用异常检测	5.8

区域代码 AL		教学大纲区域: 大数据算法	主要参考章节
层级	课题		
02	13	每个可视化技术的关键特性以及每种技术是如何使用的，并参考示例: <ol style="list-style-type: none"> 1. Bar charts – 柱形图 2. Histograms - 直方图 3. Scatter plots – 散点图 4. Bi-plots - 双标图 5. Box plots - 箱型图 6. Q-Q plots - 正态分布 Q-Q 图 7. Pie charts - 饼状图 	5.9

区域代码 PR		教学大纲范围: 大数据流程	主要参考章节
层级	课题		
记忆与大数据流程相关的关键术语 需要记忆:			
01	01	用于大数据中三个的主要流程及其主要特征	6.1
01	02	在数据分析流程中的步骤是通常使用的下列工具/技术以及它们在该步骤中的应用方式: <ol style="list-style-type: none"> 1. Data identification graph - 数据识别图 2. Data visualization techniques - 数据可视化技术 3. Algorithms - 算法 	6.2
了解大数据流程的特点、活动和技术 需要理解:			
02	01	形成大数据项目业务目标的六类问题的特点: <ol style="list-style-type: none"> 1. Descriptive - 描述性 2. Exploratory - 探索性 3. Inferential – 推理性 4. Predictive - 预测性 5. Causal - 因果性 6. Mechanistic - 机理性 	6.2

区域代码 PR		教学大纲范围: <i>大数据流程</i>	主要参考章节
02	02	数据分析流程中每个步骤的重要性以及每个步骤中发生的情况: <ol style="list-style-type: none"> 1. 确定业务目标 2. 数据识别 3. 数据收集和采购 4. 数据评审 5. 数据清洗 6. 模型建立 7. 数据处理 8. 沟通结果 	6.2
02	03	数据治理流程中每个步骤的重要性以及每个步骤中发生的情况: <ol style="list-style-type: none"> 1. 制定数据质量战略 2. 评审合规性和隐私性需求 3. 制定数据治理政策 4. 分配角色和职责 	6.3
02	04	数据管理流程中每个步骤的重要性以及每个步骤中发生的情况: <ol style="list-style-type: none"> 1. 指定指标和绩效指标 2. 监控和管理企业数据 3. 数据改进和验证 4. 对数据管理人员进行沟通和培训 	6.4

区域代码 FU		教学大纲范围: <i>大数据职能</i>	主要参考章节
层级	课题		
记忆与相关的关键术语大数据函数 需要记忆：			

区域代码 FU		教学大纲范围: 大数据职能	主要参考章节
01	01	卓越的大数据中心的五大支柱名称和每个支柱的关键特征: <ol style="list-style-type: none"> 1. Big Data Team - 大数据团队 2. Big Data Lab - 大数据实验室 3. Proof of Concepts - 概念验证 4. Agile Methodology - 敏捷方法 5. Charging Models - 计费模型 	7.2
了解卓越的大数据中心的收益，组织成功的六个因素及大数据团队的关键角色 需要理解：			
02	01	卓越的大数据中心的收益	7.2
02	02	大数据团队中关键角色的典型职责和技能： <ol style="list-style-type: none"> 1. Big Data Analyst - 大数据分析师 2. Big Data Scientist - 大数据科学家 3. Big Data Engineer - 大数据工程师 	7.3
02	03	大数据组织的六个成功因素	7.5

区域代码 AI		教学大纲范围: 人工智能	主要参考章节
层级	课题		
记忆与人工智能和大数据相关的关键定义和因素： 需要记忆：			
01	01	基于图灵测试的智能化的实用型定义	8.1
01	02	关于认知分析的关键因素： <ol style="list-style-type: none"> 1. 什么是认知分析 2. 认知分析与其他形式的分析之间差异的两个主要特征 	8.3
了解的人工智能的关键概念及其对大数据的重要性 需要理解：			

区域代码 AI		教学大纲范围: <i>人工智能</i>	主要参考章节
02	01	认知分析中理性主体的角色	8.3
02	02	人工智能的四种基本能力: 1. 自然语言处理 2. 知识表示 3. 自动推理 4. 机器学习	8.4
02	03	关于人工智能深度学习的特征: 1. 什么是深度学习 2. 主要用于何处...	8.5