



Enterprise Big Data Professional (EBDP)



Syllabus

April 2018

Introduction

The Big Data Framework provides a holistic and compressive approach for enterprises that aim to leverage the value of data in their organizations. The framework covers all the essential aspects of Big Data that are necessary to understand and analyze massive quantities of data.

The corresponding Enterprise Big Data Professional (EBDP®) qualification shows that candidates have a fundamental understanding of Big Data and are able to understand the structure and core elements of the Big Data Framework. The Enterprise Big Data Professional qualification is the first level of the Big Data Framework qualification scheme and covers fundamental knowledge and elementary concepts related to Big Data.

This syllabus covers the Enterprise Big Data Professional (EBDP®) examination. It is based on the 1st edition of the *Enterprise Big Data Professional* guide ('the Text'), which has been published on the Big Data Framework website (www.bigdataframework.org) and is available for all candidates who sit the examination.

The primary purpose of the syllabus is to provide a basis for accreditation of people involved with Big Data solutions and services. It documents the learning outcomes related to the use of Big Data Framework and describes the scope of the requirements a candidate is expected to meet to demonstrate that these learning outcomes have been achieved at each qualification level.

The target audience for this document is:

- Exam Board
- Exam Panel
- APMG Assessment Team
- Accredited Training Organizations
- EBPD Exam candidates.

This syllabus informs the design of the exams and provides accredited training organizations with a more detailed breakdown of what the exams will assess. Details on the exam structure and content are documented in the Enterprise Big Data Professional examination design.

Enterprise Big Data Professional Qualification

1.1 Purpose of the Enterprise Big Data Professional Qualification

The purpose of the Enterprise Big Data Professional (EBDP) qualification is to measure whether a candidate has sufficient knowledge and understanding of the Big Data Framework Text and whether a candidate is able to understand fundamental data processing techniques and algorithms in order to solve practical problems.

A certified Enterprise Big Data Professional has proficiency in the analysis practices and technology concepts that comprise and are featured in contemporary Big Data environments and tools.

1.2 Target Audience

This qualification is aimed at individuals who are involved in enterprise Big Data analysis, who require a working knowledge of the principles behind Big Data, who need to know the terminology used and some of the theory behind the practice. The target audience of the Enterprise Big Data Professional qualification therefore includes:

- Data Analysts
- Business Analysts
- IT Advisors
- IT Professionals

The Enterprise Big Data Professional qualification is an entry-level course for all professionals that aim to get a basic understanding of Big Data terminology, core concepts and Big Data solutions.

The course will provide an overview of existing technologies, but will not go into programming or implementation. There is no mandatory prerequisite to obtain the EBDP Qualification.

1.3 High Level Performance Definition of a Successful EBDP Candidate

The candidate who meets criteria of the High Level Performance Definition should as a minimum be able to recall, recognize and demonstrate understanding of the theories, design elements, processes and techniques as outlined in the Enterprise Big Data Professional Framework text. Additionally, a successful candidate should be able understand fundamental data processing techniques and algorithms in order to solve practical problems.

Specifically, (s)he should be able to demonstrate this understanding by being able to:

1. Explain the history and context of Big Data in comparison to traditional data analysis techniques and solutions.
2. Describe the fundamental characteristics of data analysis, big data, machine learning and artificial intelligence and the differences between these concepts.
3. Describe how to formulate a Big Data strategy that underpins the business drivers of Big Data and Big Data solutions in order to capture the value proposition of Big Data.
4. Discuss the high-level principles and design elements of contemporary Big Data architectures and explain their core benefits.
5. Explain fundamental Big Data algorithms and processing techniques in order to select the appropriate techniques to solve practical Big Data problems.
6. Identify how to apply the Big Data processes that are necessary in enterprise organizations to capture value from massive quantities of data.
7. Explain key functions, roles and competences in organizations that are necessary to capture long-term value from Big Data.
8. Understand the importance and concept of Artificial Intelligence and its relation to Big Data methods and solutions.

Assessment Model

Each learning outcome in the High Level Performance Definition requires the candidate to demonstrate specific knowledge and skills. For each learning outcome a number of learning outcome measures are identified which are evaluated in the examination, in accordance with the Examination Design, to confirm that the learning outcome has been achieved. These learning outcome measures are shown as syllabus topics and define the scope of the standard required to achieve the qualification.

A classification widely used when designing assessments for certification and education is the Bloom's Taxonomy of Educational Objectives. This classifies learning objectives into six ascending learning levels, each defining a higher degree of competencies and skills. (Bloom et al, 1956, Taxonomy of Educational Objectives).

APMG have incorporated this into a Learning Outcomes Assessment Model that is then used to develop each qualification's Assessment Model. The model provides a simple and systematic means for assessing and classifying the learning outcome measures.

This structured approach helps to ensure:

- The appropriate level is identified for a qualification
- A clear delineation in learning level content between different qualifications
- Wording is standardized and syllabi are presented consistently across APMG's qualification portfolio
- Exam questions and papers are consistent in their design.

The EBDP qualification examines at levels 1 (recall) and 2 (understand).

Big Data Framework Assessment Model				
	1. Recall	2. Understand	3. Apply	4. Analyze
APMG Learning Level Definition	<i>Remember previously learned information</i>	<i>Grasp the meaning and make sense of information</i>	<i>Use information to perform a skill or task</i>	<i>Identify whether information has been used appropriately according to the rules and guidance</i>
Generic APMG Headers <i>For introducing the learning outcome measures (topics) in the Syllabus</i>	Recall terms and key facts about concepts, principles and procedures from the reference material	Understand key facts, concepts, principles and procedures from the reference material	Apply key facts, concepts, principles and procedures to a given scenario	Differentiate between appropriate and inappropriate use of the reference material in a given scenario
Qualification Example	Recall terms and key facts about concepts, techniques, design elements and processes relating to the syllabus area	Understand the concepts, techniques, design elements and processes relating to the syllabus area	Apply particular concepts, techniques, design elements and processes relating to the syllabus area to a given scenario	Differentiate between appropriate and inappropriate use of particular concepts, techniques, design elements and processes relating to the syllabus area to a given scenario

Qualification Scope

The definition of scope for each qualification is presented in the syllabus tables at the end of this document. Each syllabus area is a unit of learning that relates to the reference material or training course module.

The following syllabus areas are identified.

Syllabus Area Code	Syllabus Area Title
CO	Big Data Concepts and Key Drivers
FR	The Big Data Framework
ST	Big Data Strategy
AR	Big Data Architecture
AL	Big Data Algorithms
PR	Big Data Processes
FU	Big Data Functions
AI	Artificial Intelligence

Syllabus Presentation

For each syllabus area the learning outcome measures are presented in order of learning level and are introduced by a standard header. There is only one header at each learning level for each syllabus area. The wording in this header is derived from the Assessment Model. Each measure is specific to a learning level.

A tick shows the scope of each examination in the respective column to the right of the topic description.

Practitioner qualification requirements are a summation of the Foundation and Practitioner learning outcome measures. All Foundation level requirements are required for Practitioner level but are assumed to have been met and are not directly assessed again, although Foundation level knowledge and understanding will be used when demonstrating Practitioner application and analysis learning outcomes.

Each of the syllabus areas is presented in a similar format as follows:

Syllabus Area Code		Syllabus Area:	Primary References
CO [2]		BD Key Concepts (CO) [1]	
Level	Topic		
Recall key terms and definitions relating to Big Data			
Specifically to recall: [3]			
01 [4]	01 [5]	[6] The definition of Big Data	[7] 1.1
01	02		

Key to the Syllabus Area table

1	Syllabus Area	Unit of learning, e.g. course module, key activity area or section of the reference guide.
2	Syllabus Area Code	A unique 2-character code identifying the syllabus area.
3	Learning Level Header	Header introducing the syllabus topics (<i>learning outcome measures</i>) for a given learning level.
4	Level	Learning level of the learning outcome measure.
5	Topic Reference	Number of the topic within the learning level.
6	Topic Description (<i>Learning Outcome Measure</i>)	Precise and specific description of what is required of the candidate to demonstrate that a learning outcome has been achieved.
7	Primary Reference	The main reference supporting the learning outcome measure.

Important Points

The following points about the use of the syllabus should be noted.

1.4 Enterprise Big Data Professional Guide References

The Enterprise Big Data Professional guide references provided should be considered to be indicative rather than comprehensive, i.e. there may be other valid references within the guidance.

Syllabus Area Code		Syllabus Area: <i>Big Data Key Concepts</i>	Primary References
CO			
Level	Topic		
Recall key terms and definitions relating to Big Data Specifically to recall:			
01	01	The definition of Big Data	1.1
01	02	The names of the four characteristics of Big Data	1.4
01	03	The names of the two classes of machine learning and the techniques commonly associated with them: 1. Supervised - classified and regression 2. Unsupervised - clustering and correlation	1.7
Understand the origins of Big Data and the characteristics of its key concepts Specifically to understand:			
02	01	The origins of Big Data and the characteristics of the three Big Data development phases: 1. Phase 1 2. Phase 2 3. Phase 3	1.3
02	02	The four characteristics of Big Data and how they distinguish Big Data from traditional data analysis: 1. Volume 2. Velocity 3. Variety 4. Veracity	1.4
02	03	The four forms of pattern identification: 1. analysis 2. analytics 3. business intelligence 4. Big Data	1.5
02	04	The purpose of the different types of analytics: 1. descriptive 2. diagnostic 3. predictive 4. prescriptive.	1.5
02	05	The function of metadata in Big Data environments	1.6
02	06	The characteristics of the three data types: 1. Structured 2. Unstructured 3. Semi-structured	1.6
02	07	The role of Hadoop in distributed storage and distributed processing	1.7

Syllabus Area Code		Syllabus Area: <i>Big Data Key Concepts</i>	Primary References
CO			
02	08	The two classes of machine learning and be able to recognize examples of these: 1. Supervised 2. Unsupervised	1.8

Syllabus Area Code		Syllabus Area: <i>The Big Data Framework</i>	Primary References
FR			
Level	Topic		
Recall terms and key facts about the Big Data Framework Specifically to recall:			
01	01	The names of the six capabilities of the Big Data Framework	2.2
Understand the structure of the Big Data Framework Specifically to understand:			
02	01	The relevance of each of the six Big Data Framework capabilities in establishing a Big Data organization	2.2
02	02	The different levels of the Big Data maturity model: <ol style="list-style-type: none"> 1. Level 1 - Analytically Impaired 2. Level 2 - Localized Analytics 3. Level 3 - Analytical Operation 4. Level 4 - Analytical Enterprise 5. Level 5 - Data Driven Enterprise 	2.4

Syllabus Area Code		Syllabus Area: <i>Big Data Strategy</i>	Primary References
ST			
Level	Topic		
Recall key facts about the Big Data Strategy Specifically to recall:			
01	01	The five steps for formulating a Big Data Strategy and their sequence	3.1
Understand how to formulate a Big Data Strategy and the activities and techniques involved Specifically to understand:			
02	01	The six business drivers influencing the need for a Big Data strategy and how Big Data can be used to generate a competitive advantage	3.1 and 3.2
02	02	The Prioritization Matrix 1. Its purpose 2. Its structure	3.3
02	03	The activities involved in each of the five steps for formulating a Big Data Strategy: 1. Step 1 - Define business objectives 2. Step 2 - Execute current state assessment 3. Step 3 - Identify and prioritize Use Cases 4. Step 4 - Formulate a Big Data Roadmap 5. Step 5 - Embed through Change Management	3.3

Syllabus Area Code		Syllabus Area: <i>Big Data Architecture</i>	Primary References
AR			
Level	Topic		
Recall terms and key facts about Big Data Architecture Specifically to recall:			
01	01	What a reference architecture is and its purpose	4.2
01	02	Key features about the structure of the NIST Big Data reference architecture: <ol style="list-style-type: none"> 1. The overall structure (5 logical roles and 2 dimensions) 2. The names of the roles 3. The names of the dimensions 4. How information flows between the different roles 	4.2
01	03	The names of the core components in a Hadoop Architecture: <ol style="list-style-type: none"> 1. NameNode 2. MapReduce 3. SlaveNode 4. Job tracker 5. HDFS 	4.6
Understand the high-level principles and design elements of contemporary Big Data Architecture Specifically to understand:			
02	01	The benefits of using a Big Data reference architecture	4.2
02	02	The functions and activities associated with the logical roles in the reference architecture <ol style="list-style-type: none"> 1. System Orchestrator 2. Data Provider 3. Big Data Application Provider 4. Big Data Framework Provider 5. Data Consumer 	4.2
02	03	The difference between local and distributed storage and processing	4.3
02	04	The three types of Big Data storage systems for massive data: <ol style="list-style-type: none"> 1. Direct Attached Storage (DAS) 2. Network Attached Storage (NAS) 3. Storage Area Network (SAN) 	4.4
02	05	The storage mechanisms for Big Data <ol style="list-style-type: none"> 1. File systems 2. NoSQL databases 3. Parallel programming models 	4.4

Syllabus Area Code AR		Syllabus Area: <i>Big Data Architecture</i>	Primary References
02	06	The Big Data Real analysis architectures: 1. Real time analysis 2. Off-line analysis	4.5
02	07	The function of Hadoop in Big Data Environments	4.6
02	08	The role of the following Hadoop components: 1. NameNode 2. MapReduce 3. SlaveNode 4. Job tracker 5. HDFS	4.6

Syllabus Area Code		Syllabus Area: <i>Big Data Algorithms</i>	Primary References
AL			
Recall terms and key facts about Big Data Algorithms and Analysis Techniques Specifically to recall:			
01	01	What descriptive statistics are	5.2
01	02	Key facts about correlation: 1. What correlation is 2. The two types of variable used in correlation 3. Key facts about the Pearson correlation coefficient: - What it measures - Its value range - What a negative, positive or 0 value means	5.4
01	03	Key facts about classification 1. What it does 2. What form of machine learning it is	5.6
Understand the algorithms and analysis techniques fundamental to Big Data Specifically to understand:			
02	01	For each type of descriptive statistic, understand what each statistical operation/distribution measures or shows: 1. Central tendency statistics 2. Dispersion statistics and 3. Distribution Shapes	5.2
02	02	The characteristics of skew: 1. Positive 2. Negative	5.2
02	03	The reason why standardization is used in Big Data calculations	5.2
02	04	Recognize and calculate examples of descriptive statistics	5.2
02	05	The characteristics of the different types of distribution shapes: 1. Frequency 2. Probability 3. Sampling 4. Normal	5.2
02	06	Why the distribution shapes are important to Big Data and data science: 1. Probability 2. Sampling 3. Normal 4. Skew	5.2
02	07	The implications of population, sample and bias for Big Data	5.3
02	08	How correlations are used in Big Data and recognize examples of this.	5.4
02	09	The differences between correlation and regression	5.5
02	10	Recognize examples of a classification algorithm	5.6

Syllabus Area Code AL		Syllabus Area: <i>Big Data Algorithms</i>	Primary References
02	11	The key characteristics of clustering: <ol style="list-style-type: none"> 1. What it does 2. Typically what most clustering algorithms look at 	5.7
02	12	How outlier detection is used in the context of Big Data	5.8
02	13	The key characteristics of each of the Visualization techniques and how each technique is used, with reference to examples: <ol style="list-style-type: none"> 1. Bar charts 2. Histograms 3. Scatter plots 4. Bi-plots 5. Box plots 6. Q-Q plots 7. Pie charts 	5.9

Syllabus Area Code		Syllabus Area:	Primary References
PR		Big Data Processes	
Level	Topic		
Recall key terms relating to the Big Data Processes			
Specifically, to recall:			
01	01	The three different main processes that are used in Big Data and their main characteristics	6.1
01	02	In which step in the data analysis process are the following tools/techniques typically used and how they are applied in that step: <ul style="list-style-type: none"> 1. Data identification graph 2. Data visualization techniques 3. Algorithms 	6.2
Understand the characteristics, activities and techniques of the Big Data Processes			
Specifically, to understand:			
02	01	The characteristics of the six types of problems that shape the business objectives of Big Data projects: <ul style="list-style-type: none"> 1. Descriptive 2. Exploratory 3. Inferential 4. Predictive 5. Causal 6. Mechanistic 	6.2
02	02	The importance of each step within the data analysis process and what occurs in each step; <ul style="list-style-type: none"> 1. Determine the business objective 2. Data identification 3. Data collection and sourcing 4. Data review 5. Data cleansing 6. Model building 7. Data processing 8. Communicating the results 	6.2
02	03	The importance of each step within the data governance process and what occurs in each step: <ul style="list-style-type: none"> 1. Develop data quality strategy 2. Review regulatory and privacy requirements 3. Develop data governance policies 4. Assign roles and responsibilities 	6.3
02	04	The importance of each activity within the data management process and the what occurs in each activity: <ul style="list-style-type: none"> 1. Specify metrics and performance indicators 2. Monitor and manage enterprise data 3. Data improvement and validation 4. Communicate and educate on data management 	6.4

Syllabus Area Code		Syllabus Area: <i>Big Data Functions</i>	Primary References
FU			
Level	Topic		
Recall key terms relating to Big Data Functions Specifically, to recall:			
01	01	The names of the five pillars of the Big Data Centre of Excellence and the key characteristics of each pillar: <ul style="list-style-type: none"> 1. Big Data Team 2. Big Data Lab 3. Proof of Concepts 4. Agile Methodology 5. Charging Models 	7.2
Understand the benefits of the Big Data Centre of Excellence, the six organization success factors and the key roles in Big Data teams Specifically, to understand:			
02	01	The benefits of a Big Data Centre of Excellence:	7.2
02	02	The typical responsibilities and skill sets of the key roles in Big Data teams: <ul style="list-style-type: none"> 1. Big Data Analyst 2. Big Data Scientist 3. Big Data Engineer 	7.3
02	03	The six organization success factors for Big Data	7.5

Syllabus Area Code AI		Syllabus Area: <i>Artificial Intelligence</i>	Primary References
Level	Topic		
Recall key definitions and facts relating to Artificial Intelligence and Big Data Specifically, to recall:			
01	01	The operational definition of intelligence according to the Turing test	8.1
01	02	Key facts about cognitive analytics: 1. What cognitive analytics is 2. The two main features that differentiate cognitive analytics from other forms of analytics	8.3
Understand the key concept of Artificial Intelligence and their importance to Big Data Specifically, to understand:			
02	01	The role of rational agents in cognitive analytics	8.3
02	02	The four essential capabilities of artificial intelligence: 1. Natural language processing 2. Knowledge representation 3. Automated reasoning 4. Machine learning	8.4
02	03	Key characteristics about Deep Learning in artificial intelligence: 1. What Deep Learning is 2. Where it is predominantly used	8.5