



**CCC
Big Data**

FOUNDATION

Certified by


Sample Exam

Edition 201606

Copyright © Cloud Credential Council 2016. All rights reserved.
Copyright © EXIN Holding B.V. 2016. All rights reserved.
EXIN® is a registered trademark.

No part of this publication may be reproduced, stored, utilized or transmitted in any form or by any means, electronic, mechanical, or otherwise, without the prior written permission from EXIN.

Content

| | |
|--------------|----|
| Introduction | 4 |
| Sample exam | 5 |
| Answer key | 13 |
| Evaluation | 27 |

Introduction

This is the sample exam EXIN CCC Big Data Foundation. The Rules and Regulations for EXIN's examinations apply to this exam.

This exam consists of 40 multiple-choice questions. Each multiple-choice question has a number of possible answers, of which only one is the correct answer.

The maximum number of points that can be obtained for this exam is 40. Each correct answer is worth one point. If you obtain 26 points or more you will pass.

The time allowed for this exam is 60 minutes.

Good luck!

Sample exam

1 / 40

Which of the following is **not** a term for measuring data quantity?

- A. Alphabyte
- B. Gigabyte
- C. Kilobyte
- D. Zettabyte

2 / 40

One of the early articles about Big Data discusses volume, velocity and variety.

What is the name of the author of this article?

- A. Doug Laggett
- B. Doug Laney
- C. Doug Langley
- D. Derek Laney

3 / 40

Who should be involved in a Big Data project?

- A.
 - C-Suite members
 - Domain Experts
 - Data Administrators
 - Project Managers
- B.
 - C-Suite members
 - Domain Experts
 - Scary Data People
 - IT professionals
- C.
 - Data Administrator
 - IT professionals
 - Project Managers
 - Scary Data People
- D.
 - Domain Experts
 - IT professionals
 - Project Managers
 - Scary Data People

4 / 40

According to McKinsey Global Institute, how much money can potentially be saved in the US Health Care system when using Big Data?

- A. \$150 million
- B. \$300 million
- C. \$500 million
- D. \$300 billion

5 / 40

According to the International Maritime Bureau (IMB), piracy decreases as a result of using Big Data.

According to IMB, what is the decrease in piracy in the first 6 months of 2012?

- A. 12.5%
- B. 21%
- C. 54%
- D. 80%

6 / 40

What percentage of people in the US can be identified by just date-of-birth, ZIP-code, and gender?

- A. 7.6%
- B. 16%
- C. 45%
- D. 85%

7 / 40

Which company uses MongoDB® as the basis of its content management system?

- A. Associated Press
- B. BBC
- C. Forbes
- D. Time Warner

8 / 40

eBay is running a Hadoop® cluster.

How many nodes does their cluster have?

- A. 79
- B. 121
- C. 532
- D. 812

9 / 40

What are the three main characteristics of Big Data?

- A. Validity, Velocity, Veracity
- B. Velocity, Variety, Volume
- C. Volume, Veracity, Variety
- D. Volume, Validity, Venality

10 / 40

What are the names of the main players in the Enterprise Systems space?

- A. Facebook, Instagram, Twitter
- B. MS[®] Access, MySQL[™], DB2
- C. PeopleSoft, JD Edwards, Quora
- D. SAP[®], Oracle[®], Microsoft[®]

11 / 40

Which Facebook API provides information only about user's friends?

- A. Graph API
- B. Keyword Inserts API
- C. Public Feed API
- D. Rest API

12 / 40

What are data warehouses, **as a rule** based on?

- A. Hierarchical database technology
- B. Network database technology
- C. Relational database technology
- D. Spreadsheet files

13 / 40

What is the **main** interest, when working with metadata?

- A. Information, such as, date and time, author, and origin
- B. Message content
- C. Recipient's personal details
- D. Time when a database table was created

14 / 40

To whom is the Public Feed API of Facebook available?

- A. A limited set of media publishers
- B. A select number of Silicon Valley start-ups
- C. Some universities in US and Canada
- D. The general public

15 / 40

Which datasets are focused on providing data linked to various government and inter-government policies?

- A. Public Feed API
- B. Rest API
- C. Search API
- D. Streaming API

16 / 40

Which datasets are focused on providing data linked to various government and inter-government policies?

- A. Economic databases
- B. Financial databases
- C. Monetary databases
- D. Governmental databases

17 / 40

Which streaming APIs does Twitter provide?

- A. Daily streams, monthly streams, weekly streams
- B. Facebook streams, Google streams, user streams
- C. Public streams, data streams, user streams
- D. Public streams, user streams, site streams

18 / 40

Which type of APIs is used by **applications** to create tweets, read author profiles, and follow information?

- A. Graph APIs
- B. Public Feed APIs
- C. REST APIs
- D. Streaming APIs

19 / 40

What type of process is Data Mining?

- A. Discovering patterns within large datasets.
- B. Extracting data from datasets.
- C. Maintaining large datasets by removing and adding records.
- D. Updating data elements in large datasets.

20 / 40

What is **not** a type of Data Mining?

- A. Association
- B. Classification
- C. Clustering
- D. Cubing

21 / 40

What is the main algorithm used for Clustering?

- A. K-averages
- B. K-means
- C. L-averages
- D. L-means

22 / 40

Weka is a popular data mining application.

Who developed this application?

- A. The University of Chicago
- B. The University of Konstanz
- C. The University of Queensland
- D. The University of Waikato

23 / 40

What are two **main** Hadoop® modules?

- A. HDFS™ and YARN™
- B. MapReduce and NoSQL
- C. Pig and SQL
- D. RDBMS and NoSQL

24 / 40

Where is the failover mechanism of Hadoop® implemented?

- A. Data centers
- B. Hardware
- C. Servers
- D. Software

25 / 40

What is **true** about DataNodes and/or NameNodes?

- A. DataNode periodically sends a HeartBeat to a NameNode.
- B. DataNode is a server managing access to files and the namespace.
- C. NameNode manages the storage on a node.
- D. NameNode periodically sends a HeartBeat to a DataNode.

26 / 40

Which type of Hadoop® installation is useful for debugging and testing?

- A. Local
- B. Pseudo-distributed
- C. Fully-distributed
- D. Virtual

27 / 40

Which of the following is **not** a type of Hadoop® installation?

- A. Local
- B. Fully-distributed
- C. Pseudo-distributed
- D. Hybrid

28 / 40

What is **true** about MapReduce™?

- A. Is used to map relational tables to Hadoop®.
- B. Takes several datasets and combines them for further processing.
- C. Takes a big dataset and splits it across several nodes.
- D. Uses similar syntax to SQL.

29 / 40

Which Hadoop® component is **not** suitable on clusters with less than five nodes?

- A. MapReduce™
- B. Hadoop® Common
- C. HBase™
- D. Hive™

30 / 40

What type of model is used for the YARN™ resource management component?

- A. Master-slave model
- B. Master-master model
- C. Slave-slave model
- D. Peer-to-peer model

31 / 40

Which of the following is **not** provided by Hive™?

- A. Access to HDFS™ files
- B. Extract/Transform/Load (ETL) functionality
- C. SQL functionality
- D. Query Language (QL) functionality

32 / 40

What type of database is MongoDB®?

- A. Document database
- B. MapReduce™ database
- C. Open source relational database
- D. Relational database

33 / 40

Which feature of MongoDB® allows closer approximation of real-world entities as compared to other document databases?

- A. Document-oriented architecture
- B. Embedded structures
- C. Replication
- D. Sharding

34 / 40

For which operating system is MongoDB® not available?

- A. Chrome OS™
- B. Mac OS®
- C. Linux™
- D. Solaris™

35 / 40

What type of scheme is used for document databases?

- A. Bi-directional
- B. Denormalized
- C. Fixed
- D. Flexible

36 / 40

In a document database, what does a **collection** correspond to?

- A. A record in a relational database
- B. A field in a relational database
- C. A key-value pair in Hadoop®
- D. A table in a relational database

37 / 40

What is the maximum size of a document in MongoDB®?

- A. 8 MB
- B. 16 MB
- C. 128 MB
- D. 1 GB

38 / 40

With what can names in a MongoDB® **not** start?

- A. A number
- B. % character
- C. \$ character
- D. # character

39 / 40

Which MongoDB® package contains the daemon and relevant configuration and init scripts?

- A. mongodb-org
- B. mongodb-org-server
- C. mongodb-org-shell
- D. mongodb-org-tools

40 / 40

Which MongoDB® package contains mongos daemon?

- A. mongodb-org
- B. mongodb-org-mongos
- C. mongodb-org-server
- D. mongodb-org shell

Answer key

1 / 40

Which of the following is **not** a term for measuring data quantity?

- A. Alphabyte
- B. Gigabyte
- C. Kilobyte
- D. Zettabyte

- A. Correct. Kilobyte, Gigabyte, and Zettabyte are used for measuring data quantity.
- B. Incorrect
- C. Incorrect
- D. Incorrect

2 / 40

One of the early articles about Big Data discusses volume, velocity and variety.

What is the name of the author of this article?

- A. Doug Laggett
- B. Doug Laney
- C. Doug Langley
- D. Derek Laney

- A. Incorrect
- B. Correct. Doug Laney published a paper on February 6th, 2001 on Big Data discussing volume, velocity, and variety.
- C. Incorrect
- D. Incorrect

3 / 40

Who should be involved in a Big Data project?

A.

- C-Suite members
- Domain Experts
- Data Administrators
- Project Managers

B.

- C-Suite members
- Domain Experts
- Scary Data People
- IT professionals

C.

- Data Administrator
- IT professionals
- Project Managers
- Scary Data People

D.

- Domain Experts
- IT professionals
- Project Managers
- Scary Data People

A. Incorrect

B. Correct. Normally, C-Suite members, Domain Experts, Scary Data People, and IT professionals are involved in a Big Data project.

C. Incorrect

D. Incorrect

4 / 40

According to McKinsey Global Institute, how much money can potentially be saved in the US Health Care system when using Big Data?

A. \$150 million

B. \$300 million

C. \$500 million

D. \$300 billion

A. Incorrect

B. Incorrect

C. Incorrect

D. Correct. According to McKinsey Global Institute (MGI), the potential benefit of Big Data for US Health Care is close to \$300 billion.

5 / 40

According to the International Maritime Bureau (IMB), piracy decreases as a result of using Big Data.

According to IMB, what is the decrease in piracy in the first 6 months of 2012?

- A. 12.5%
- B. 21%
- C. 54%
- D. 80%

- A. Incorrect
- B. Incorrect
- C. Correct. The International Maritime Bureau reports that there is a fall of 54% in piracy in the first 6 months of 2012, because of the use of Big Data.
- D. Incorrect

6 / 40

What percentage of people in the US can be identified by just date-of-birth, ZIP-code, and gender?

- A. 7.6%
- B. 16%
- C. 45%
- D. 85%

- A. Incorrect
- B. Incorrect
- C. Incorrect
- D. Correct. 85% of people in US can be identified using DOB, ZIP, and gender combined.

7 / 40

Which company uses MongoDB® as the basis of its content management system?

- A. Associated Press
- B. BBC
- C. Forbes
- D. Time Warner

- A. Incorrect
- B. Incorrect
- C. Correct. Forbes uses MongoDB® as the basis of its content management system.
- D. Incorrect

8 / 40

eBay is running a Hadoop® cluster.

How many nodes does their cluster have?

- A. 79
- B. 121
- C. 532
- D. 812

- A. Incorrect
- B. Incorrect
- C. Correct. eBay is running a Hadoop® cluster of 532 nodes (8*532, 5.3 PB).
- D. Incorrect

9 / 40

What are the three main characteristics of Big Data?

- A. Validity, Velocity, Veracity
- B. Velocity, Variety, Volume
- C. Volume, Veracity, Variety
- D. Volume, Validity, Venality

- A. Incorrect
- B. Correct. The 3 main characteristics of Big Data are:
 - Velocity
 - Variety
 - Volume
- C. Incorrect
- D. Incorrect

10 / 40

What are the names of the main players in the Enterprise Systems space?

- A. Facebook, Instagram, Twitter
- B. MS® Access, MySQL™, DB2
- C. PeopleSoft, JD Edwards, Quora
- D. SAP®, Oracle®, Microsoft®

- A. Incorrect
- B. Incorrect
- C. Incorrect
- D. Correct. The main players in the Enterprise System Space are Oracle®, SAP®, and Microsoft®.

11 / 40

Which Facebook API provides information only about user's friends?

- A. Graph API
 - B. Keyword Inserts API
 - C. Public Feed API
 - D. Rest API
- A. Correct. The Graph API provides information only about user's friends and not about everyone on Facebook.
- B. Incorrect
- C. Incorrect
- D. Incorrect

12 / 40

What are data warehouses, **as a rule** based on?

- A. Hierarchical database technology
 - B. Network database technology
 - C. Relational database technology
 - D. Spreadsheet files
- A. Incorrect
- B. Incorrect
- C. Correct. A data warehouse is normally a relational database that is designed for query and analysis, rather than for transaction processing.
- D. Incorrect

13 / 40

What is the **main** interest, when working with metadata?

- A. Information, such as, date and time, author, and origin
 - B. Message content
 - C. Recipient's personal details
 - D. Time when a database table was created
- A. Correct. Metadata is defined as "data about data" and may include:
- time of creating or sending an email
 - the author or sender
 - the recipient
 - the quantity of files or emails
- B. Incorrect
- C. Incorrect
- D. Incorrect

14 / 40

To whom is the Public Feed API of Facebook available?

- A. A limited set of media publishers
- B. A select number of Silicon Valley start-ups
- C. Some universities in US and Canada
- D. The general public

- A. Correct. The Public Feed API of Facebook provides access to users and page status updates. This API is available to a limited set of media publishers.
- B. Incorrect
- C. Incorrect
- D. Incorrect

15 / 40

Which datasets are focused on providing data linked to various government and inter-government policies?

- A. Public Feed API
- B. Rest API
- C. Search API
- D. Streaming API

- A. Correct. The Public Feed API provides data linked to various government and inter-government policies.
- B. Incorrect
- C. Incorrect
- D. Incorrect

16 / 40

Which datasets are focused on providing data linked to various government and inter-government policies?

- A. Economic databases
- B. Financial databases
- C. Monetary databases
- D. Governmental databases

- A. Correct. Economics datasets are focused on providing data linked to various government and inter-government policies and manifested in a number of economic indicators.
- B. Incorrect
- C. Incorrect
- D. Incorrect

17 / 40

Which streaming APIs does Twitter provide?

- A. Daily streams, monthly streams, weekly streams
 - B. Facebook streams, Google streams, user streams
 - C. Public streams, data streams, user streams
 - D. Public streams, user streams, site streams
-
- A. Incorrect
 - B. Incorrect
 - C. Incorrect
 - D. Correct. The 3 main streaming APIs provided by Twitter are:
 - public streams
 - user streams
 - site streams

18 / 40

Which type of APIs is used by **applications** to create tweets, read author profiles, and follow information?

- A. Graph APIs
 - B. Public Feed APIs
 - C. REST APIs
 - D. Streaming APIs
-
- A. Incorrect
 - B. Incorrect
 - C. Correct. Using REST APIs, applications can create tweets, read author profiles and follow information.
 - D. Incorrect

19 / 40

What type of process is Data Mining?

- A. Discovering patterns within large datasets.
 - B. Extracting data from datasets.
 - C. Maintaining large datasets by removing and adding records.
 - D. Updating data elements in large datasets.
-
- A. Correct. Data mining is the process of discovering patterns within large datasets.
 - B. Incorrect
 - C. Incorrect
 - D. Incorrect

20 / 40

What is **not** a type of Data Mining?

- A. Association
 - B. Classification
 - C. Clustering
 - D. Cubing
-
- A. Incorrect
 - B. Incorrect
 - C. Incorrect
 - D. Correct. The three main types of data mining are:
 - Classification
 - Clustering
 - Association

21 / 40

What is the main algorithm used for Clustering?

- A. K-averages
 - B. K-means
 - C. L-averages
 - D. L-means
-
- A. Incorrect
 - B. Correct. K-means is the main algorithm used for clustering. With K-means we decide how many clusters are needed. This is represented by the value of K.
 - C. Incorrect
 - D. Incorrect

22 / 40

Weka is a popular data mining application.

Who developed this application?

- A. The University of Chicago
 - B. The University of Konstanz
 - C. The University of Queensland
 - D. The University of Waikato
-
- A. Incorrect
 - B. Incorrect
 - C. Incorrect
 - D. Correct. Weka is a popular data mining application, written in Java, and developed at The University of Waikato, New Zealand. The name comes from **Waikato Environment for Knowledge Analysis**.

23 / 40

What are two **main** Hadoop® modules?

- A. HDFS™ and YARN™
- B. MapReduce and NoSQL
- C. Pig and SQL
- D. RDBMS and NoSQL

- A. Correct. HDFS and YARN are two main types of Hadoop modules.
- B. Incorrect
- C. Incorrect
- D. Incorrect

24 / 40

Where is the failover mechanism of Hadoop® implemented?

- A. Data centers
- B. Hardware
- C. Servers
- D. Software

- A. Incorrect
- B. Incorrect
- C. Incorrect
- D. Correct. The failover mechanism of Hadoop® is implemented in the software, further reducing the cost of implementation.

25 / 40

What is **true** about DataNodes and/or NameNodes?

- A. DataNode periodically sends a HeartBeat to a NameNode.
- B. DataNode is a server managing access to files and the namespace.
- C. NameNode manages the storage on a node.
- D. NameNode periodically sends a HeartBeat to a DataNode.

- A. Correct. A DataNode periodically sends a HeartBeat to a NameNode indicating that they're functioning properly.
- B. Incorrect
- C. Incorrect
- D. Incorrect

26 / 40

Which type of Hadoop® installation is useful for debugging and testing?

- A. Local
- B. Pseudo-distributed
- C. Fully-distributed
- D. Virtual

- A. Correct. The Local (standalone) installation is useful for debugging and testing.
- B. Incorrect
- C. Incorrect
- D. Incorrect

27 / 40

Which of the following is **not** a type of Hadoop® installation?

- A. Local
 - B. Fully-distributed
 - C. Pseudo-distributed
 - D. Hybrid
-
- A. Incorrect
 - B. Incorrect
 - C. Incorrect
 - D. Correct. Hadoop can be installed through the following three types of installations:
 - Local
 - Pseudo-distributed
 - Fully-distributed

28 / 40

What is **true** about MapReduce™?

- A. Is used to map relational tables to Hadoop®.
 - B. Takes several datasets and combines them for further processing.
 - C. Takes a big dataset and splits it across several nodes.
 - D. Uses similar syntax to SQL.
-
- A. Incorrect
 - B. Incorrect
 - C. Correct. MapReduce™ works by taking a big dataset and splitting it amongst several nodes to process.
 - D. Incorrect

29 / 40

Which Hadoop® component is **not** suitable on clusters with less than five nodes?

- A. MapReduce™
 - B. Hadoop® Common
 - C. HBase™
 - D. Hive™
-
- A. Incorrect
 - B. Incorrect
 - C. Correct. HBase™ is best applied when tables with hundreds of millions or billions of records are present and enough hardware is present. However, on clusters with less than five nodes, not much can be achieved through HBase™.
 - D. Incorrect

30 / 40

What type of model is used for the YARN™ resource management component?

- A. Master-slave model
 - B. Master-master model
 - C. Slave-slave model
 - D. Peer-to-peer model
- A. Correct. YARN™ is a Hadoop® component responsible for resource management and job scheduling, which is set up in a Master-Slave configuration. The Master receives the requests; the slaves handle them.
- B. Incorrect
 - C. Incorrect
 - D. Incorrect

31 / 40

Which of the following is **not** provided by Hive™?

- A. Access to HDFS™ files
 - B. Extract/Transform/Load (ETL) functionality
 - C. SQL functionality
 - D. Query Language (QL) functionality
- A. Incorrect.
- B. Incorrect.
- C. Correct. You cannot use SQL within Hive. Hive uses a query language called HiveQL.
- D. Incorrect.

32 / 40

What type of database is MongoDB®?

- A. Document database
 - B. MapReduce™ database
 - C. Open source relational database
 - D. Relational database
- A. Correct. MongoDB® is a Document database.
- B. Incorrect
 - C. Incorrect
 - D. Incorrect

33 / 40

Which feature of MongoDB® allows closer approximation of real-world entities as compared to other document databases?

- A. Document-oriented architecture
- B. Embedded structures
- C. Replication
- D. Sharding

A. Correct. The Document-oriented architecture feature of MongoDB® allows closer approximation of real-world entities as compared to other document databases.

- B. Incorrect
- C. Incorrect
- D. Incorrect

34 / 40

For which operating system is MongoDB® not available?

- A. Chrome OS™
- B. Mac OS®
- C. Linux™
- D. Solaris™

A. Correct. MongoDB is available for Mac OS X®, Microsoft Windows®, Solaris™, and Linux™.

- B. Incorrect
- C. Incorrect
- D. Incorrect

35 / 40

What type of scheme is used for document databases?

- A. Bi-directional
- B. Denormalized
- C. Fixed
- D. Flexible

- A. Incorrect
- B. Incorrect
- C. Incorrect

D. Correct. Document databases, including MongoDB, have a flexible schema.

36 / 40

In a document database, what does a **collection** correspond to?

- A. A record in a relational database
- B. A field in a relational database
- C. A key-value pair in Hadoop®
- D. A table in a relational database

- A. Incorrect
- B. Incorrect
- C. Incorrect
- D. Correct. In a document database, a collection stores a number of documents. It is the equivalent of a table in a relational database.

37 / 40

What is the maximum size of a document in MongoDB®?

- A. 8 MB
- B. 16 MB
- C. 128 MB
- D. 1 GB

- A. Incorrect
- B. Correct: In MongoDB® the maximum size of a document is 16 MB.
- C. Incorrect
- D. Incorrect

38 / 40

With what can names in a MongoDB® **not** start?

- A. A number
- B. % character
- C. \$ character
- D. # character

- A. Incorrect
- B. Incorrect
- C. Correct. Field names are strings and must adhere to certain restrictions. Names in MongoDB® cannot start with the \$ character.
- D. Incorrect

39 / 40

Which MongoDB® package contains the daemon and relevant configuration and init scripts?

- A. mongodb-org
 - B. mongodb-org-server
 - C. mongodb-org-shell
 - D. mongodb-org-tools
-
- A. Incorrect
 - B. Correct. The mongodb-org-server package contains the daemon and relevant configuration and init scripts.
 - C. Incorrect
 - D. Incorrect

40 / 40

Which MongoDB® package contains mongos daemon?

- A. mongodb-org
 - B. mongodb-org-mongos
 - C. mongodb-org-server
 - D. mongodb-org shell
-
- A. Incorrect
 - B. Correct. The mongodb-org-mongos package contains mongos daemon (shard routing services).
 - C. Incorrect
 - D. Incorrect

Evaluation

The table below shows the correct answers to the questions in this sample exam.

| Question | Answer | Question | Answer |
|----------|--------|----------|--------|
| 1 | A | 21 | B |
| 2 | B | 22 | D |
| 3 | B | 23 | A |
| 4 | D | 24 | D |
| 5 | C | 25 | A |
| 6 | D | 26 | A |
| 7 | C | 27 | D |
| 8 | C | 28 | C |
| 9 | B | 29 | C |
| 10 | D | 30 | A |
| 11 | A | 31 | C |
| 12 | C | 32 | A |
| 13 | A | 33 | A |
| 14 | A | 34 | A |
| 15 | A | 35 | D |
| 16 | A | 36 | D |
| 17 | D | 37 | B |
| 18 | C | 38 | C |
| 19 | A | 39 | B |
| 20 | D | 40 | B |

Contact EXIN

www.exin.com

