Home / I'm Learning / BigData$ IoT / Chapter 6: Architecture for Big Data and Data Engineering / Chapter 6 Quiz

Course Home

BigData$ IoT

Grades

Started on	Tuesday, 6 June 2023, 4:00 PM
State	Finished
Completed on	Tuesday, 6 June 2023, 4:50 PM
Time taken	49 mins 48 secs
Marks	29.00/34.00
Grade	**85.29** out of 100.00

Messages

Calendar

Question **1**

Incorrect

Mark 0.00 out of 2.00

Which statement describes the term containers in virtualization technology?

Select one:

○ a group of VMs with identical OS and applications

◉ a subsection of a virtualization environment that contains one or more VMs   ✖

○ a virtual area with multiple independent applications sharing the host OS and hardware

○ isolated areas of a virtualization environment, where each area is administered by a customer

Refer to curriculum topic: 6.1.2
In a virtualization environment, containers are a specialized "virtual area" where multiple applications can run independently of each other while sharing the same OS and hardware. By sharing the host operating system, most of the software resources are reused, which leads to reduced boot time and optimized operation.

The correct answer is: a virtual area with multiple independent applications sharing the host OS and hardware

Question **2**

Correct

Mark 2.00 out of 2.00

Which file system does Hadoop use?

Select one:

○ FAT32

◉ HDFS   ✔

○ NTFS

○ SAS

Refer to curriculum topic: 6.2.3
Hadoop uses the Hadoop Distributed File System (HDFS) which is made for servers in a data storage cluster. HDFS has the ability to create one large file system from multiple servers to provide increased storage, performance, and redundancy.

The correct answer is: HDFS

Question **3**

Correct

Mark 2.00 out of 2.00

Which architecture is used by HDFS?

Select one:

○ peer-to-peer

○ master-slave   ✔

○ client/server

○ stand alone

Refer to curriculum topic: 6.3.2
Hadoop Distributed File System (HDFS) employs a master-slave architecture, whereas Cassandra File System (CFS) employs a peer-to-peer
implementation.

The correct answer is: master-slave

Question **4**

Correct

Mark 2.00 out of 2.00

Which transferring methodology do traditional message brokers use?

Select one:

○ publish-subscribe   ✔

○ transaction logging

○ client-server

○ real-time

Refer to curriculum topic: 6.3.1
Traditionally, transferring messages consists of two different methods:

- Publish-Subscribe – The requested messages are broadcast to all of the consumers.
- Point-to-Point – Multiple consumers read messages from the server. Each of these messages goes to one of the consumers.

The correct answer is: publish-subscribe

Question **5**

Correct

Mark 2.00 out of 2.00

Which scenario is suitable for fog computing?

Select one:

○ Multiple sensors are used to collect national climate information.

○ Sales data of a national department store are collected and analyzed.

○ Multiple motion sensors are implemented at major street intersections in a city.   ✔

○ Orders on a major online store are processed and fulfilled from multiple warehouse locations.

Refer to curriculum topic: 6.1.1
Information collected at a street intersection needs immediate processing in order for officials to make decisions for quick adjustment of traffic lights.
Thus fog computing should be implemented. Orders on an online store are sent to a centralized data center to process and to locate an appropriate
warehouse to ship orders. This is true for collecting and analyzing sales data and national client-patterns. Climate information collected from multiple
sensors distributed in a wide area usually does not need an immediate response, the information should be sent to cloud computing for aggregation
and analysis.

The correct answer is: Multiple motion sensors are implemented at major street intersections in a city.

Question **6**

Correct

Mark 2.00 out of 2.00

Which function is provided by MapReduce to Hadoop?

Select one:

○ storage expansion

○ hypervisor security

◉ distributed processing   ✔

○ database management

Refer to curriculum topic: 6.2.2
Developed by Google, MapReduce is also used in the Hadoop Big Data ecosystem for distributed processing to manage scalability.

The correct answer is: distributed processing

Question **7**

Correct

Mark 2.00 out of 2.00

Which type of hypervisor would most likely be used in a data center?

Select one:

◉ Type 1   ✔

○ Type 2

○ Nexus

○ Hadoop

Refer to curriculum topic: 6.1.2
The two type of hypervisors are Type 1 and Type 2. Type 1 hypervisors are usually used on enterprise servers. Enterprise servers rather than virtualized PCs are more likely to be in a data center.

The correct answer is: Type 1

Question **8**

Correct

Mark 2.00 out of 2.00

Which solution improves web response time by deploying multiple web servers and DNS servers?

Select one:

○ sharding

○ memcaching

◉ load balancing   ✔

○ distributed databases

Refer to curriculum topic: 6.2.2
Maintaining availability is the primary concern for companies working with big data. Some solutions to improve the availability include the following:

- **Load Balancing** – deploying multiple web servers and DNS servers to respond to requests simultaneously
- **Distributed Databases** – improving database access speed and demands
- **Memcaching** – offloading demand on database servers by keeping frequently requested data available in memory for fast access
- **Sharding** – partitioning a large relational database across multiple servers to improve search speed

The correct answer is: load balancing

Question 9

Correct

Mark 2.00 out of 2.00

What are two benefits to an organization if it rented data center services from either a co-location facility or a cloud service provider? (Choose two.)

Select one or more:

The organization will reduce the WAN connection cost.

The internal IT staff will get training on data center operations.

Rented data centers are usually close to the organization for easy access.

✓ The organization pays a smaller up-front cost and requires fewer internal IT staff. ✔

✓ Rented data centers and equipment are managed by highly experienced professionals 24/7. ✔

Refer to curriculum topic: 6.1.1
There are a few benefits for an organization to rent data center service at co-location facilities or from cloud service providers, including the following:

- **Security** - Secure space is used to house equipment and electronically-controlled access. Security staff are available 24/7.
- **Flexibility** - Additional capacity can be added quickly and cheaply.
- **Redundancy / Backup** - Redundant and backup systems are built into the system design.
- **Location** - Data centers are implemented to be away from high-probability disaster areas.
- **Management** - Facilities and equipment are managed by experienced professionals 24/7.
- **High Return on investment** - Organizations leasing space have smaller up-front costs, and fewer internal IT staff are required.

The organization will actually increase usage of WAN connections, and the internal operation of those data centers will not be open to customers.

The correct answers are: The organization pays a smaller up-front cost and requires fewer internal IT staff., Rented data centers and equipment are managed by highly experienced professionals 24/7.

Question 10

Correct

Mark 2.00 out of 2.00

What is the purpose of device virtualization?

Select one:

○ It allows the physical layer to interface with another device through multiple virtual connections.

○ It allows multiple physical PCs to be linked as one virtual PC.

◉ It allows multiple OSs to run on one physical device.   ✔

○ It allows a networking device to manage multiple virtual local area networks.

Refer to curriculum topic: 6.1.2
Device virtualization consolidates the number of required devices by allowing multiple operating systems to exist on a single hardware platform.

The correct answer is: It allows multiple OSs to run on one physical device.

Question 11

Correct

Mark 2.00 out of 2.00

Match the cloud computing service to the description. (Not all options are used.)

network equipment	IaaS	✔
development tools	PaaS	✔
virtual desktop environment	SaaS	✔

The correct answer is: network equipment → IaaS, development tools → PaaS, virtual desktop environment → SaaS

Question 12

Correct

Mark 2.00 out of 2.00

A cloud service provider is considering building a data center on the coast of southern California. They have found a pre-existing building with large windows facing the ocean. What are two factors the provider should consider? (Choose two.)

Select one or more:

political instability

site elevation

likelihood of natural disasters ✓

security ✓

the number of homes in the area

Refer to curriculum topic: 6.1.1
Factors to consider when building a data center include location, electrical requirements, environmental, security, and network design. When considering the location, the planning team should consider areas with reduced risk of natural disasters such as earthquakes, floods, and fire. Data centers are also usually located away from high traffic areas including airports and malls, and areas of strategic importance to governments and utilities such as refineries, dams, and nuclear reactors. Data centers are located in cities where corporations would take advantages of their services.

The correct answers are: likelihood of natural disasters, security

Question 13

Correct

Mark 2.00 out of 2.00

What is hyperjacking?

Select one:

○ taking over a virtual machine hypervisor as part of a data center attack ✓

○ overclocking the mesh network which connects the data center servers

○ using multiple virtual machines as one large Hadoop server running HDFS

○ using processors from multiple computers to increase data processing power

Refer to curriculum topic: 6.1.1
Hyperjacking occurs when an attacker hijacks a virtual machine (VM) hypervisor and then uses that VM to launch an attack on other data center devices.

The correct answer is: taking over a virtual machine hypervisor as part of a data center attack

Question **14**

Incorrect

Mark 0.00 out of 2.00

What are two data storage problems with big data? (Choose two.)

Select one or more:

- ✓ data management   ✓
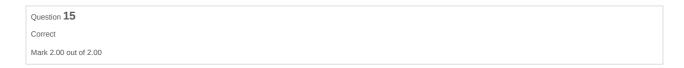- increased maintenance cost
- unstructured data
- ✓ application performance   ✗
- loss of predictive analysis

Refer to curriculum topic: 6.3.2

There are at least five data storage problems with big data:

- **Management** – the existence of thousands of data management tools are available but with very few data-sharing standards.
- **Security** – Authentication, access, and accounting is difficult to ensure.
- **Unstructured Data** – Unstructured data is difficult to search and analyze.
- **Input and Output** – Large projects that produce extremely large data sets tax storage solutions with the amount of I/O requests that they make.
- **The WAN** – As more storage moves to the cloud, larger levels of strain is placed on WAN links.

The correct answers are: data management, unstructured data

Question **15**

Correct

Mark 2.00 out of 2.00

Match the job in a data center to the description.

data scientist ✓	Take raw data and turn it into meaningful information. Apply statistics, machine learning, and analytic approaches to answer critical business
business analyst ✓	Study a business or an industry and then formulate a specific question. Work with company stakeholders to determine the issue of concern.
data analyst ✓	Query and process data, provide reports, and summarize and visualize data.
data engineer ✓	Create the infrastructure that supports big data. Design and build the platform on which all of this data is stored and processed.

Refer to curriculum topic: 6.2.1

The correct answer is: data scientist → Take raw data and turn it into meaningful information. Apply statistics, machine learning, and analytic approaches to answer critical business questions., business analyst → Study a business or an industry and then formulate a specific question. Work with company stakeholders to determine the issue of concern., data analyst → Query and process data, provide reports, and summarize and visualize data., data engineer → Create the infrastructure that supports big data. Design and build the platform on which all of this data is stored and processed.

Question **16**

Correct

Mark 2.00 out of 2.00

Why can it be more beneficial to use Spark than MapReduce when creating a big data solution?

Select one:

- ⭘ batch processing
- 🔵 built-in machine learning library   ✔
- ⭘ data stored on disk
- ⭘ parallelizing algorithms

Refer to curriculum topic: 6.3.3

Spark is gaining popularity because of its performance, ease of administration, simplicity and the fact that applications can be created more quickly using it. Some of the differentiating features are as follows:

- It is capable of dealing with enormous amounts of real-time data.
- It can transcend different silos of data.
- It supports many different languages, which means there is less code that needs to be written and maintained.
- It is easier to learn to develop and less intimidating than MapReduce.

The correct answer is: built-in machine learning library

Question **17**

Partially correct

Mark 1.00 out of 2.00

Match the big data tool to the description.

HDFS	open source distributed data processing engine	✖
Cassandra	open-source NoSQL distributed database management system	✔
Spark	a system that uses a master-slave architecture	✖
Kafka	messaging system that uses transaction logs	✔

The correct answer is: HDFS → a system that uses a master-slave architecture, Cassandra → open-source NoSQL distributed database management system, Spark → open source distributed data processing engine, Kafka → messaging system that uses transaction logs

◄ Chapter 6 Terms and Concepts Practice

Jump to...

EoC Survey ►

Terms and Conditions

Privacy Statement

Cookie Policy

Data Protection

Trademarks

Data Protection

Accessibility