📄 Quiz name
bigdata final 2022

📅 Date played
Oct 21, 2022 1:13 PM

👤 Hosted by
Dr. Chantana CHANTRAPORNCHAI

Accuracy	
84%	◎

Class accuracy	
56%	👥

Graded Questions		Correct		Incorrect		Avg. time taken	
50	❓	**42**	✓	**8**	✕	**29.4s**	⏱

Replay this quiz with your child

Play now

Find similar quizzes

Detailed Report
≔ 50 questions

Question 1 Correct	⏱ 44 secs 🪙 1 points

Question Type: Multiple-choice

What is correct about RDD ? (Select all that apply)

Papon CHOONHAKLAI's Response

It supports data partitioning and distributed through cluster.	data frame can be converted to RDD with python supported.	CSV and JSON can be accepted.

Correct answer

It supports data partitioning and distributed through cluster.	data frame can be converted to RDD with python supported.	CSV and JSON can be accepted.

Question 2 Incorrect	⏱ 26 secs 🪙 0 points

Question Type: Fill-in-the-Blank

Two types of RDD operations are ...., ...

reduce,count

Correct answer

transformation and action

Question 3   Correct   ⏱ 25 secs   💰 1 points

Question Type: Multiple-choice

What is the RDD operation that gives outputs? (select all that apply)

Papon CHOONHAKLAI's Response

Reduce   Count

Correct answer

Reduce   Count

Question 4   Incorrect   ⏱ 39 secs   💰 0 points

Question Type: Multiple-choice

What is true about RDD Persistence? (Select all that apply)

Papon CHOONHAKLAI's Response

It caches results in memory only as it it in-memory processing.

The automatic recomputation is done if RDD partition is lost.

Correct answer

It has cache to speedup future operations.

The automatic recomputation is done if RDD partition is lost.

Question 5   Correct   ⏱ 43 secs   💰 1 points

Question Type: Multiple-choice

What is true about Spark and Hadoop mapreduce? (Select all that apply)

Papon CHOONHAKLAI's Response

Spark has faster performance due to its in-memory processing feature.

Spark can run on top of HDFS or not.

Correct answer

Spark has faster performance due to its in-memory processing feature.

Spark can run on top of HDFS or not.

Question 6   Correct   ⏱ 12 secs   💰 1 points

Question Type: Multiple-choice

when using jupyter notebook, the command
ssh -N -L p1:localhost:p2 hadoop@10.3.133.231 what is p1, p2?

p1 is local port

p2 is remote machine port

Correct answer

p1 is local port

p2 is remote machine port

Question 7   Correct   ⏱ 39 secs   🪙 1 points

Question Type: Multiple-choice

what cannot we do in one command for graphframe? (Select all that apply)

Papon CHOONHAKLAI's Response

find subgraph   find depthfirst search

Correct answer

find subgraph   find depthfirst search

Question 8   Correct   ⏱ 37 secs   🪙 1 points

Question Type: Multiple-choice

Kafka broker has functions: (Select all that apply)

Papon CHOONHAKLAI's Response

message exchange center   Monitor the incoming message   buffer the message

Correct answer

message exchange center   Monitor the incoming message   buffer the message

---

Question 9   Correct   ⏱ 21 secs   🪙 1 points

Question Type: Multiple-choice

what is the important element we need when running kafka?

Papon CHOONHAKLAI's Response

zookeeper

Correct answer

zookeeper

Question 10   Correct   ⏱ 7 secs   🪙 1 points

Question Type: Multiple-choice

what is the default port for kafka broker?

Papon CHOONHAKLAI's Response

9092

Correct answer

Question 11   Correct   ⏱ 39 secs   🪙 1 points

Question Type: Multiple-choice

What is correct about data storage in HBase? (select all that apply)

Papon CHOONHAKLAI's Response

Table is divided by row key.   Rows are split into region.

Correct answer

Table is divided by row key.   Rows are split into region.

Question 12   Correct   ⏱ 45 secs   🪙 1 points

Question Type: Multiple-choice

What is correct about the design criteria for HBase? (Select all that apply)

Papon CHOONHAKLAI's Response

The redundancy is allowed to increase parallelism.   Empty column will not be stored.

Correct answer

The redundancy is allowed to increase parallelism.   Empty column will not be stored.

Question 13   Incorrect   ⏱ 14 secs   🪙 0 points

Question Type: Multiple-choice

Hive uses ....for execution engine.

Papon CHOONHAKLAI's Response

Zookeeper

Correct answer

Ted

Question 14   Correct   ⏱ 33 secs   🪙 1 points

Question Type: Multiple-choice

HiveServer2 is .... (select all that apply)

Papon CHOONHAKLAI's Response

for handing multiuser session.   connected through jdbc/odbc

Correct answer

for handing multiuser session.   connected through jdbc/odbc

Question 15   Correct   ⏱ 27 secs   🪙 1 points

Question Type: Multiple-choice

What is correct about Pig? (Select all that apply)

Papon CHOONHAKLAI's Response

It is a kind of data-flow language.

The script is transformed into several Mapreduce tasks.

Correct answer

It is a kind of data-flow language.

The script is transformed into several Mapreduce tasks.

Question 16   Correct   ⏱ 18 secs   💰 1 points

Question Type: Multiple-choice

The following command in Pig triggers mapreduce job. (Select all that apply)

Papon CHOONHAKLAI's Response

dump   store

Correct answer

dump   store

Question 17   Correct   ⏱ 33 secs   💰 1 points

Question Type: Multiple-choice

What are not important elements in hbase-site.xml?

Papon CHOONHAKLAI's Response

dfs replication

Correct answer

dfs replication

Question 18   Correct   ⏱ 9 secs   💰 1 points

Question Type: Fill-in-the-Blank

what is the default port for kibana?

Papon CHOONHAKLAI's Response

5601

Correct answer

5601

Question 19   Correct   ⏱ 45 secs   💰 1 points

Question Type: Multiple-choice

what is correct about elasticsearch? (Select all that apply)

Papon CHOONHAKLAI's Response

it supports the full text   Index is equivalent to   we query the document based

search. document on json.

Correct answer

it supports the full text search.

Index is equivalent to document

we query the document based on json.

Question 20   Correct   ⏱ 36 secs   💰 1 points

Question Type: Multiple-choice

what is correct about logstash ? (select all that apply)

Papon CHOONHAKLAI's Response

it can handle multiple data source.

we use logstash to filter data before entering elasticsearch to reduce the datasize

one filebeat instance can handle multiple files.

Correct answer

it can handle multiple data source.

we use logstash to filter data before entering elasticsearch to reduce the datasize

one filebeat instance can handle multiple files.

Question 21   Correct   ⏱ 35 secs   💰 1 points

Question Type: Multiple-choice

Logstash can output to more than just elasticsearch.

Papon CHOONHAKLAI's Response

True

Correct answer

True

Question 22   Correct   ⏱ 35 secs   💰 1 points

Question Type: Fill-in-the-Blank

what is the keyword for output of logstash to output to screen?

Papon CHOONHAKLAI's Response

rubydebug

Correct answer

stdout
rubydebug

Question 23   Correct   ⏱ 24 secs   💰 5 points

Question Type: Multiple-choice

which command can be used to test the existence of Elasticsearch process on terminal?

Papon CHOONHAKLAI's Response

curl

Correct answer

curl

Question 24   Correct   ⏱ 27 secs   💰 5 points

Question Type: Multiple-choice

if elasticsearch gets killed during the starting up, what is the likely cause? (select all that apply)

Papon CHOONHAKLAI's Response

jvm heap size is too large for available memory.

Correct answer

jvm heap size is too large for available memory.

Question 25   Correct   ⏱ 17 secs   💰 5 points

Question Type: Multiple-choice

logstash can accept ....number of input sources.

Papon CHOONHAKLAI's Response

multiple

Correct answer

multiple

Question 26   Correct   ⏱ 12 secs   💰 5 points

Question Type: Multiple-choice

.................is a feature for logstash used to filter incoming data.

Papon CHOONHAKLAI's Response

grok

Correct answer

grok

Question 27   Correct   ⏱ 33 secs   💰 5 points

Question Type: Multiple-choice

index is equivalent to .... in mysql.

Papon CHOONHAKLAI's Response

database

Correct answer

database

Question 28   Correct   ⏱ 30 secs   🪙 5 points

Question Type: Multiple-choice

what is advantage of Parquet over pandas? (select all that apply)

Papon CHOONHAKLAI's Response

data compression   have embedded schema

Correct answer

data compression   have embedded schema

Question 29   Correct   ⏱ 55 secs   🪙 5 points

Question Type: Multiple-choice

Apache parquet ……………………… (select all that apply)

Papon CHOONHAKLAI's Response

is a column data   is stored in a key-value   can be converted from
store   format.   pandas

Correct answer

is a column data   is stored in a key-value   can be converted from
store   format.   pandas

Question 30   Correct   ⏱ 7 secs   🪙 5 points

Question Type: Multiple-choice

Apache parquet data store is saved in a single file.

Papon CHOONHAKLAI's Response

False

Correct answer

False

Question 31   Incorrect   ⏱ 28 secs   🪙 0 points

Question Type: Multiple-choice

What is advantage of dask over pandas? (select all that apply)

Papon CHOONHAKLAI's Response

can run on GPU   can run on cluster   can read any size data frame.

Correct answer

can read multiple files at the   can run on   can run on   can read any size data
same time.   GPU   cluster   frame.

Question 32   Correct   ⏱ 25 secs   🪙 5 points

Question Type: Multiple-choice

result = df.attribute1.mean()

what is the next line in order to show the

value of mean of attribute1?

Papon CHOONHAKLAI's Response

result.compute()

Correct answer

result.compute()

Question 33   Correct   ⏱ 47 secs   🪙 5 points

Question Type: Multiple-choice

when the user browses the visualization and plays interactive, what affects the response time to the
user? select all that apply.

Papon CHOONHAKLAI's Response

server CPU speed   bus speed between memory and cpu   internet bandwidth

Correct answer

server CPU speed   bus speed between memory and cpu   internet bandwidth

Question 34   Incorrect   ⏱ 32 secs   🪙 0 points

Question Type: Multiple-choice

what is scheduled into each SM?

Papon CHOONHAKLAI's Response

thread

Correct answer

thread block

Question 35   Correct   ⏱ 22 secs   🪙 5 points

Question Type: Multiple-choice

Increasing more cores always increases performance.

Papon CHOONHAKLAI's Response

False

Correct answer

False

Question 36   Incorrect   ⏱ 16 secs   🪙 0 points

Question Type: Multiple-choice

GPU centric data movement tries to .... (select all that apply)

Move data between GPUs among nodes via RDMA

Move data between host and device

Correct answer

Move data between GPUs among nodes via RDMA

Move data between GPUs in a host using NVLink

Question 37   Correct   ⏱ 51 secs   💰 5 points

Question Type: Multiple-choice

What is pinned memory?

Papon CHOONHAKLAI's Response

Unpageable memory

Correct answer

Unpageable memory

Question 38   Correct   ⏱ 28 secs   💰 5 points

Question Type: Multiple-choice

NVTabular dataset supports...........................(Select all that apply)

Papon CHOONHAKLAI's Response

Dask   Pandas DataFrame   Parquet file

Correct answer

Dask   Pandas DataFrame   Parquet file

Question 39   Correct   ⏱ 3 secs   💰 5 points

Question Type: Multiple-choice

NVTabular can read multiple files at one command.

Papon CHOONHAKLAI's Response

True

Correct answer

True

Question 40   Correct   ⏱ 36 secs   💰 5 points

Question Type: Multiple-choice

To initialize workflow, parameters needed are ....(select all that apply)

Papon CHOONHAKLAI's Response

column name   categorial columns   continuous value columns

Correct answer

column name   categorial columns   continuous value columns

Question 41   Correct   31 secs   5 points

Question Type: Multiple-choice

What are the types of map operations? select all that apply.

Papon CHOONHAKLAI's Response

x=x+1   read (x)

Correct answer

x=x+1   read (x)

Question 42   Incorrect   51 secs   0 points

Question Type: Multiple-choice

What are the ops in NVTablular for feature engineering and preprocessing ? select all that apply.

Papon CHOONHAKLAI's Response

normalize   zerofill

Correct answer

normalize   log   zerofill

Question 43   Correct   50 secs   5 points

Question Type: Multiple-choice

In which case, parquet is faster than csv format. Select all that apply.

Papon CHOONHAKLAI's Response

select one column

Correct answer

select one column

Question 44   Correct   23 secs   5 points

Question Type: Multiple-choice

In which situation, parquet cudf is slower than csv cudf.

Papon CHOONHAKLAI's Response

skips_row=20,col_indexes = range(0, 7000)

Correct answer

skips_row=20,col_indexes = range(0, 7000)

Question 45   Correct   20 secs   5 points

Question Type: Multiple-choice

In the visualization of DLI, what is the output of the callback? Select all that apply.

Papon CHOONHAKLAI's Response

Graph

Correct answer

Graph

Question 46   Correct   ⏱ 35 secs   🪙 5 points

Question Type: Multiple-choice

what is correct about the assessment? select all that apply.

Papon CHOONHAKLAI's Response

| need to convert to pandas df before display on dash figure | need to drop na value | set na_value to dummy -9999 |

Correct answer

| need to convert to pandas df before display on dash figure | need to drop na value | set na_value to dummy -9999 |

Question 47   Correct   ⏱ 28 secs   🪙 5 points

Question Type: Reorder

What is the typical process for workflow pipeline? order it.

Papon CHOONHAKLAI's Response

add new features

fillmissing

normalize

Correct answer

add new features
fillmissing
normalize

Question 48   Incorrect   ⏱ 32 secs   🪙 0 points

Question Type: Multiple-choice

What is needed for mqtt installation? select all that apply

Papon CHOONHAKLAI's Response

mosquitto-clients   mosquitto_pub

Correct answer

mosquitto   mosquitto-clients

Question 49   Correct   ⏱ 25 secs   🪙 1 points

Question Type: Multiple-choice

what is correct about MQTT? select all that apply.

Papon CHOONHAKLAI's Response

default port is 1883   '#' is used as wildcard multilevel,

Correct answer

default port is 1883   '#' is used as wildcard multilevel,

Question 50   Correct   🕐 16 secs   🪙 1 points

Question Type: Multiple-choice

MQTT can do.................... and .... while Kafka model is producer/ consumer

Papon CHOONHAKLAI's Response

publish/subscribe

Correct answer

publish/subscribe