HPCCSystems\_ConnectionsHomes

**Welcome hackers!**

This quiz is part of your final submission. The more questions you answer correctly the more points you will receive during judging session. This quiz is designed to help you learn some quick facts about big data and ECL (Section 1) and emphasizes some important items about the challenge (Section 2).

Since the answers cannot be saved here, a word version of quiz is available in the Git repository. Use the word version to save your answers. Once you are satisfied with your answers transfer your answers to this quiz. Remember we need only ONE submission per team. Thank you.

General Questions

1. What is your team's name?
2. Including yourself, add all your team members' names and (optional) their email addresses.
3. The term big data refers to:
   1. Structured data
   2. Any data that can be stored in a computer
   3. Extremely large data sets
   4. None of the above
4. What is Metadata? (1 point)
   1. Data preview
   2. Information that describes properties and context of data
   3. Data layout
   4. Rules on how to display data
5. What kind of data does the following define?

Data can be processed, analyzed, stored and retrieved in a fixed (defined) format. It has both form and format.

1. Unstructured
2. Variable
3. Structured
4. Data processing
5. None of the above
6. Data Mapping is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
   1. True
   2. False
7. What does ETL in big data stand for?
   1. Extract, Tolerate, Language
   2. Extract, Transform, Load
   3. Extract, Transform, Load
   4. Execution, Tolerate, Load
8. Which of the following are part of Data Life Cycle? Select all that apply.
   1. Data delivery
   2. Data collection
   3. Data collection
   4. Data collection
   5. Data remodeling
9. Data structure is about data management, storage format, retrieving and processing.
   1. True
   2. False
10. The value of big data is in the collecting of the data.
    1. True
    2. False
11. The following is an example of Unstructured Data:
12. Video file
13. Spreadsheet
14. Address book
15. None of the above
16. What are the 3 Vs in data?
    1. Volume, Veracity, Value
    2. Value, Variety, Velocity
    3. Variability, Visualization, Volume
    4. Volume, Variety, Velocity
17. Data privacy is not important during data collection phase.
    1. True
    2. False
18. Parallel processing is about dividing the task among multiple workers who then operate on their part simultaneously.
    1. True
    2. False
19. Which of the following is an example of Structured Data?
20. PDF document
21. Word document
22. Spreadsheet file
23. All of the above
24. Which of the followings manages large amount of data at once and has the least time sensitivity?
    1. Stream Processing
    2. Real-time processing
    3. Batch processing
    4. Data processing
25. When submitting an ECL job, into which language is it converted for compiling?
26. HPCC
27. C++
28. SQL
29. Java
30. SHARED is usable outside MODULE.
31. True
32. False
33. What is the result of following code?

Text

Description automatically generated

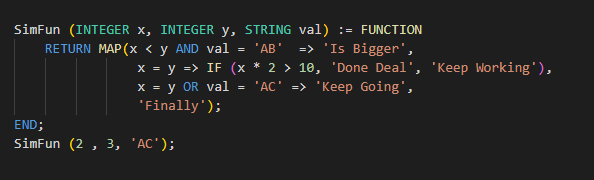
* 1. Red, 200
  2. White, 300
  3. 200, 300
  4. 300

1. Which of the following specifies exactly how each field in the output record set is to receive its value?
2. JOIN
3. PROJECT
4. TRANSFORM
5. RECORD
6. What are the keywords for XXX and YYY?

Text

Description automatically generated

1. SELF, RIGHT
2. LEFT, RIGHT
3. SELF, LEFT
4. LEFT, SELF
5. What is the result of following?



1. Finally
2. Keep Going
3. Keep Working
4. Done Deal
5. How many rows the following code will generate?

Text

Description automatically generated

* 1. 1
  2. 2
  3. 3
  4. 4

1. What is the output of following?

STD.Str.CountWords('Today//sun/ /is/shining/', '/');

1. Today sun is shining,
2. Today sun is shining,
3. 4
4. 5
5. 6
6. In a JOIN if we want to find only non-matched records from the left dataset, which of the following flags do we need?
   1. LEFT OUTER
   2. LEFT ONLY
   3. RIGHT OUTER
   4. RIGHT ONLY
   5. FULL OUTER
7. How many rows following code will generate?

Text

Description automatically generated

1. 0
2. 1
3. 2
4. 3
5. What is the result of the following code?

AVE(TRUNCATE(3), TRUNCATE(2));

1. 2
2. 2.5
3. 5
4. 3
5. ECL is a case sensitive language
   1. True
   2. False
6. Is following code correct?

UNSIGNED2 Func (STRING val, INTEGER num) := num \* 2;

1. Yes
2. No
3. What is the default JOIN type?
4. LEFT ONLY
5. INNER
6. FULL
7. None of the above
8. Is the following valid?

DEDUP(SORT(DISTRIBUTE(allPeople, HASH32(personID)), RECORD, LOCAL), RECORD , LOCAL);

1. True
2. False
3. JOIN can have one dataset.
4. True
5. False
6. What is the output of following?

Text

Description automatically generated

1. 0
2. 1
3. True
4. False
5. What is the output of following?

val := 5;

CHOOSE(val, 'Some', 'Today', 'Book');

1. Some
2. Today
3. Book
4. None of the above

Connections Homes Questions

1. What is the purpose of our matching algorithm?
   1. Help mentors
   2. Help youth find home
   3. All of the above
   4. None of the above
2. What is the minimum required input?
   1. Youth name
   2. Distance
   3. Religious
   4. All of the above
3. When do we need Human Review Flag?
   1. No information in a category is provided
   2. Half fields are filled in
   3. Distance is not provided
   4. All of the above
4. If a mentor has 0 weight for a value, that mentor is eliminated.
   1. True
   2. False
5. How many youths should compare against all mentors at a time?
   1. All
   2. 1
   3. 2
   4. 50%
6. What interface should be used to capture youth information?
7. Coder choice
8. Roxie query
9. Thor
10. None of the above
11. Should youth dataset be used in the solution?
12. Yes
13. No
14. Do I need to be working with all fields in mentor dataset?
15. Yes, all fields are required for matching
16. No, only the fields that relative to matching solution and are available on both sides (youth and mentor)
17. What is the first step to solve this problem?
18. Understanding the data and what is available to us
19. Working on solution
20. Testing plan
21. None of the above
22. When working with data the first step is to filter the data down to what is needed
23. True
24. False
25. Using plain English explain the mapping solution you using to match a category in youth to all its related information in mentor data set? For example, mapping Alcohol Use input to Occasionally, Irresponsibly, Responsibility, or N/A in mentor’s dataset. (Text box answer use words and not code)
26. Using plain English explain how you calculate your distance? (Text box answer use words and not code)
27. Using plain English explain how your matching algorithm? (Text box answer use words and not code)
28. What is your final WUID? Make sure the WUID includes your complete and final solution.  Judges will be using this WUID to evaluate your code, solution and your result. This WUID should include your active Roxie query as well.
29. What hurdles did you have to overcome to complete this challenge?
30. What did you learn from this challenge? All answers count as correct!
31. If you were supposed to solve this problem again, what would you differently? All answers count as correct!
32. Including yourself, how many of your team will attend the next HPCC event? All answers count as correct!
33. Is any of your team member interested in our internship program? If yes, please leave their name and contact information.

Internship program: <https://wiki.hpccsystems.com/display/hpcc/HPCC+Systems+Summer+Intern+Program>  
All answers count as correct!

1. We are always looking for ways to improve our events any feedback is appreciated. All answers count as correct!