70-247: Web and Distributed Programming
Final Exam Information

Exam Format

The final exam will be taken in-class, during finals week, through the BlackBoard system. You will have 90 minutes to complete the exam. **Anyone coming in late will not be allowed to take the exam and will receive a zero score.** You will not be able to use your book, notes, phones, and other devices, and cannot run any software on the lab computers except the browser for running BlackBoard. You can however bring two sheets of empty paper and pencils to use for scratch work.

The exam will consist of 50 questions, 45 of which will come from the same pool of questions as quizzes and midterms and 5 will be new free response questions. This means that you will not receive an automatic score for this exam until I grade them. Once all the exams are graded, the scores will then be normalized as stated in the syllabus.

Topics

All material covered by the midterm exams 1 and 2 plus PHP material not covered in the second midterm, databases, using PHP with databases, cloud computing, and Hadoop/mapreduce, including:

- PHP
    - Basic data types: integer, double, string, boolean
    - Using basic functions: rand, abs, trim, …
    - Output: using print/echo. Know how and when value substitution is performed.
    - Know the difference between client and server scripting
    - Control statements: if-else, switch statement, while, for, do-while loops
    - Arrays: initialization, accessing elements, iterating through elements using iterators or foreach loop, sorting.
    - Functions: passing parameters, scope
    - Matching patterns (preg_match, preg_split functions)
    - Handling forms: accessing form data (_GET or _POST arrays)
    - Saving/accessing cookies
    - Using sessions
- Databases
    - Structure of relational databases
    - Making simple SQL queries: select, insert, delete, …
    - Accessing MySQL from PHP
        - Connection
        - Query
        - Accessing result
- Cloud Computing
    - Definition
    - Characteristics
    - Models
    - Advantages
    - Applications
- Hadoop/MapReduce

- HDFS
  - architecture: how does HDFS store/access files
  - what are advantages to using a Cloud File System like HDFS?
- MapReduce
  - What defines a MapReduce program?
  - architecture: how are MapReduce programs executed?
  - What are advantages of using MapReduce
  - Constructing Map and Reduce functions for a given problem

Study Questions

Below, you will find review questions that you can use to make sure you are prepared for the exam. You should review these plus all the questions that were on the midterms and quizzes. Note however, that this is just a sample and the exam questions may refer to any of the above listed topics.

Q1. Name 2 superglobal variables available in PHP.

Q2. Why is it necessary to revalidate user's form inputs after they have been validated on the client side with JavaScript?

Q3. What is the output of the following PHP scripts?

```php
<?php
$a = 5;
$b = 6;
function mult($b) {
        global $a;
        return $a * $b;
}
echo "$a*$b=" . mult(5);
?>
```

Q4. Assume there is an html page with a form that contains a textbox with attribute name set to "pwd" and a submit button. The action of the form submits the form using the GET method to a PHP script. Write a PHP script that processes this form by outputting "wrong password" if the textbox input does not equal "password".

Q5. Assume there is an html page with a form that contains a textbox with attribute name set to 'id' and a submit button. The action of the form submits the form using the POST method to a PHP script. Write a PHP script that processes this form by outputting "OK" if the textbox input includes a lower case letter followed by a sequence of digits, followed by a single upper case letter.

Q6. Assume there is an html page with a form that contains a checkbox with attribute name set to "opt1" and a value of "1", and a submit button. The action of the form submits the form using the GET method to a PHP script. Write a PHP script that starts a new session and if the checkbox with name "opt1" was checked, saves "OK" into a session variable named "state".

Q7. (True/False) The _POST and _GET arrays are accessible inside PHP functions.

Q8. Name 1 advantage of cookies of sessions.

Q9. Write the call to a PHP function that opens a file for appending.

Q10.

a. Write an SQL query that creates a new table called Students with the following fields: id, studentid, name, age.

b. Assume the table from (a) exists. Write an SQL query that returns the student id's of all students who are over 21 years old.

c. Assume the table from (a) exists. Write an SQL query that inserts a single student record into the table.

Q11. What is a relational database?

Q12. What happens when a DataNode fails in a Hadoop File System?

Q13. (True/False) Using MapReduce cluster for computation is always faster than using a single machine.

Q14. Name at least three benefits of using Cloud Computing.

ANSWERS

1. GET and POST

2. PHP can validate the input securely.

3. 5*6=25

4.

```php
<?php
if ($_GET['pwd'] != "password") {
  echo "wrong password";
}

?>
```

5.

```php
<?php

if (preg_match("/^[a-z][0-9]+[A-Z]$/", $_GET['pwd'])) {
  echo "OK";
}

?>
```

6.
```php
<?php
if (isset($_GET['opt1'])) {
   $_SESSION['state'] = 'OK';
}
?>
```

7. TRUE

8. Can persist over longer durations.

9. $file = fopen("filename", "a");

10.

a. CREATE TABLE Students
    (id INT UNSIGNED NOT NULL AUTO_INCREMENT PRIMARY KEY,
     studentid INT UNSIGNED, name char(20), age INT UNSIGNED);

b. select * from students where age>21;

c. insert into students (studentid, name, age) values (123, 'john', 21);

11. Database composed of a collection of tables, possibly related through (usually key) values.

12. The DataNode will stop sending heartbeat messages to the NameNode and therefore the NameNode will know not to send it any IO requests. If the replication level will then become lower than specified minimum, the NameNode will request a replication to occur.

13. False. It's only faster for specific applications that involve large data sets.

14. Cost savings, scalability, and reliability