

Midterm Examination

Wednesday May 16, 4:40pm to 6pm

Your name:

Instructions: Look through the whole exam and answer the questions that you find easiest first. Answer each question in the space below the question, using the backs of the pages for extra space as necessary. There are 50 points in total.

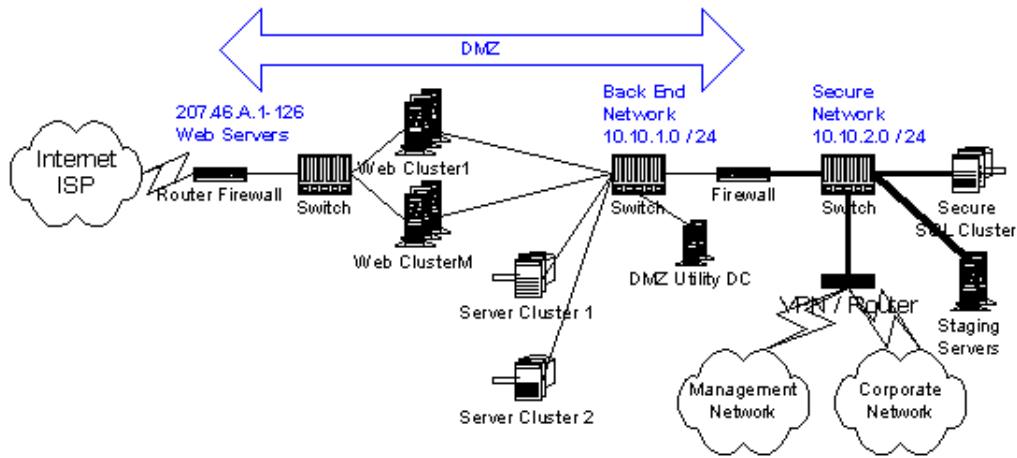
Whenever you make an assumption, state it clearly. In questions where you are asked to discuss an issue, it is important that you use the relevant concepts and terminology covered in class and in the textbooks.

You may use a calculator, one book on PHP, one book on databases, the published 134A lecture notes, and your own personal notes. No other books or notes may be used.

(Question 1) [20 points] This question refers to the figure on the next page. Write your answers below the figure. Each answer should include a short explanation (typically one or two sentences).

- (a) [3 points] What does the acronym DMZ stand for? What is the DMZ?
- (b) [3 points] Why are there two firewalls? What does each firewall do? What firewall functionality (if any) should be implemented in the router that connects to the corporate network?
- (c) [4 points] Assume that each web cluster is stateless, while each server cluster is a relational database system. For which type of cluster is it more difficult to achieve fault tolerance? Explain briefly how fault tolerance can be achieved for each type.
- (d) [2 points] What role do “staging servers” play in typical web site architectures?
- (e) [5 points] For each of the following types of data, explain where it should be stored. Justify each answer in one sentence.
 - 1. The session identifier of a client.
 - 2. Session data for a client, e.g. shopping cart contents.
 - 3. Sensitive data for a client, e.g. home telephone number.
 - 4. Public data such as product descriptions.
 - 5. Log data for the web servers.
- (f) [3 points] This architecture for a web site has at least one “single point of failure.” Identify this point and explain briefly.

Figure 1: Example web site architecture for Question 1



(Question 2) [16 points] This question refers to a table named `president` with schema as follows:

Field	Type	Null	Key	Default	Extra
<code>last_name</code>	<code>varchar(15)</code>				
<code>first_name</code>	<code>varchar(15)</code>				
<code>state</code>	<code>char(2)</code>				
<code>city</code>	<code>varchar(20)</code>				
<code>birth</code>	<code>date</code>			<code>0000-00-00</code>	
<code>death</code>	<code>date</code>	YES		NULL	

- (a) [2 points] Write an SQL query that returns the names of presidents who are still alive.
- (b) [3 points] Write an SQL query that finds the state in which the greatest number of presidents have been born.
- (c) [2 points] Can you write a query that finds the states in which no presidents have been born? Explain your answer.
- (d) [3 points] Write an SQL query that finds which presidents have the same last name.
- (e) [3 points] Why is the following query likely to be slow, regardless of what indexes are available?

```
select * from president where 2001 - year(birth) < 75
```

Explain how to write a different query that computes the same result but is much faster.

- (f) [3 points] When you are testing the speed of an SQL query with and without an index, it is important to run each version of the query several times. Explain why.

(Question 3) [14 points] Definition: An identifier is a series of mixed letters and digits, of length one or more, where the first character is a letter. Assume that identifiers are as long as possible, and therefore one identifier cannot overlap with another.

Your answer to each of the following parts should be a PHP function that is efficient and understandable.

(a) [4 points] Carefully write a PHP function using regular expressions that returns the first identifier present anywhere in a long string.

(b) [4 points] Carefully write a PHP function that returns all the identifiers in a string.

(c) [6 points] Carefully write a PHP function using regular expressions that returns the last identifier in a string. (For efficiency, this function must not be based on extracting all the identifiers in the string.)