Yelp.com interview questions

February 29, 2012

What's a HTTP GET/POST? HTTP = HyperText Transfer Protocal. GET/POST are two of the most common methods of requests. POST request is typically used to submit forms. A GET request sends its parameters via the URL (e.g.: http://robozzle.com/puzzle.aspx?id=85). A POST request sends its parameters in the request body, just under the headers.

How many bits in a short? how many bits of an integer? What is the maximum value of an unsigned integer?

Language	Type	Sign	Bytes	Bits	Word	Min	Max
C/C++	short	signed	2	16	0.5	-2^{15}	$2^{15}-1$
	unsigned short	unsigned	2	16	0.5	0	$2^{16}-1$
	int	signed	4	32	1	-2^{31}	$2^{31}-1$
	uint32_t	unsigned	4	32	1	0	$2^{32}-1$
Java	short	signed	2	16	0.5	-2^{15}	$2^{15}-1$
	int	signed	4	32	1	-2^{32}	$2^{32}-1$

What's a static variable? In computer programming, a static variable is a variable that has been allocated statically — whose lifetime extends across the entire run of the program.

What is SSL? Short for Secure Sockets Layer, a protocol developed by Netscape for transmitting private documents via the Internet. SSL uses a cryptographic system that uses two keys to encrypt data — a public key known to everyone and a private or secret key known only to the recipient of the message. Both Netscape Navigator and Internet Explorer support SSL, and many Web sites use the protocol to obtain confidential user information, such as credit card numbers. By convention, URLs that require an SSL connection start with *https:* instead of *http*.

Difference between a process and a thread? In computing, a process is an instance of a computer program that is being executed. It contains the program code and its current activity. Depending on the operating system

(OS), a process may be made up of multiple threads of execution that execute instructions concurrently.

Threads differ from traditional multitasking operating system processes in that:

- processes are typically independent, while threads exist as subsets of a process

- processes carry considerably more state information than threads, whereas multiple threads within a process share process state as well as memory and other resources

- processes have separate address spaces, whereas threads share their address space

- processes interact only through system-provided inter-process communication mechanisms

- Context switching between threads in the same process is typically faster than context switching between processes.

What are the different synchronization mechanisms?   In computer science, synchronization refers to one of two distinct but related concepts: synchronization of processes, and synchronization of data.

- synchronization of processes: most common, mutex and semaphore

- synchronization of data: RAID

Difference between semaphore and mutex?   *Mutex* is a key to a toilet. One person can have the key - occupy the toilet - at the time. When finished, the person gives (frees) the key to the next person in the queue. *Semaphore* is the number of free identical toilet keys. Example, say we have four toilets with identical locks and keys. The semaphore count - the count of keys - is set to 4 at beginning (all four toilets are free), then the count value is decremented as people are coming in. If all toilets are full, ie. there are no free keys left, the semaphore count is 0. Now, when eq. one person leaves the toilet, semaphore is increased to 1 (one free key), and given to the next person in the queue.

What is the time complexity of searching in a linked list? $O(n)$.

Given a file with a word on each line, get the top-ten most frequent words.   Use hashtabl to hash each word then use a priority queue to store and retrieve the top-ten most frequent words.

How many bytes does a 32 bit operating system work with? 2^{32} bits / 8 bits per byte $= 2^{29}$ bytes $= 536, 870, 912$ bytes. (536 millions, 870 thousands and 912).

What command would you use to retrieve one row of data from a table using SQL? SELECT ... FROM ... WHERE ... ORDER BY ... FETCH FIRST n ROWS ONLY

What is the command for viewing memory management information in Linux? One of the best ways to see both RAM and CPU usage is to type '\$ vmstat', which will display 5 sections of useful memory information, including RAM and CPU information. You can also type '\$ top', which displays a very exhaustive list of memory information, but it is easy to get the important data lost in the flood.

What is the best and worst case time complexities for a hash table? Best case: $O(1)$, worst case: $O(n)$.

Use linked list or array for binary search? Array because it's easier for random access.

What is the linux command for counting number of words in a file? wc [-clmw] [file]. -c: number of bytes. -l: number of lines. -m: number of characters. -w: number of words.

Explain process after you click a URL (on server / client side) You type in the URL and hit go. The browser needs to translate that URL www.somesite.com into an IP address so it knows what computer on the internet to connect to (That URL is just there to make it easier for us humans - kinda like speed-dial for phone numbers I guess). So your browser will see if it already has the appropriate IP address cached away from previous visits to the site. If not, it will make a DNS query to your DNS server (might be your router or your ISP's DNS server). Once your browser knows what IP to use, it will connect to the appropriate webserver and ask for the page. The webserver then returns the requested page and your browser renders it to the screen.

Explain DNS resolution The resolver, or another DNS server acting recursively on behalf of the resolver, negotiates use of recursive service using bits in the query headers. Resolving usually entails iterating through several name servers to find the needed information. The process entails:

1. A network host is configured with an initial cache (so called hints) of the known addresses of the root nameservers. Such a hint file is updated periodically by an administrator from a reliable source.

2. A query to one of the root servers to find the server authoritative for the top-level domain.

3. A query to the obtained TLD server for the address of a DNS server authoritative for the second-level domain.

4. Repetition of the previous step to process each domain name label in sequence, until the final step which returns the IP address of the host sought.

Some question based on complexities of sorting algorithm , Usual sorting algorithm complexity.

	Bubble	Merge	Quick	Insertion	Selection
Best	n	$n \log n$	$n \log n$	n	n^2
Average	n^2	$n \log n$	$n \log n$	n^2	n^2
Worst	n^2	$n \log n$	n^2	n^2	n^2
Memory	1	n	$\log n$	1	1
Stable	Yes	Yes	No	Yes	No
In-place	Yes	No	Yes	Yes	Yes

Given a Number to guess b/w 1-100 , how to point to the number in minimum time   Binary search. Minimum time: $\lceil \log_2 100 \rceil = 7$

How cookies are working?   A cookie is a piece of text that a Web server can store on a user's hard disk. Cookies allow a Web site to store information on a user's machine and later retrieve it. The pieces of information are stored as name-value pairs. In the broadest sense, a cookie allows a site to store state information on your machine. This information lets a Web site remember what state your browser is in.

The Difference Between HTTP and HTTPS?   "HTTPS" stands for "Hyper Text Transfer Protocol Secure." It means that information exchanged between you and a web site is encrypted and cannot be hijacked by someone who might want to electronically eavesdrop when you type a credit card number, a password, a social security number, or any other person information. HTTPS usually uses the SSL encryption methods.

How you can know from which website your user is coming?   HTTP headers.

Which operation is more expensive? Multiply, Divide or Add?   Divide and multiply are all more expensive than add. Depend on different implementation of CPU, ususally divide is a bit more expensive.

What is the average disk access time?   Access time = Seek time + Rotation Delay + Transfer time. An HDD's Average Access Time is its average Seek time which technically is the time to do all possible seeks divided by the number of all possible seeks.

If memory is limited, which data structure is better: hashtable or binary tree? Binary trees use memory in proportion to the data that is to be stored - the data and links. For a hash table, if the hash function has a huge range, you will have to allocate an array that has an entry for each number in the range. You could potentially store far lesser number of elements than the range.

Given a telephone directory which is broken in number of files, how would you find a telephone number in it. Use "grep" command in Linux and search for regex.