Speed up your web pages

The first of its kind Web Technology Conference on Open Source Technology, WebOSS '07 was being organised in Kolkata today and spoke at the event as one of the participants and thought I will share my presentation here. As download requires you to register at the forum I am sharing the content of the presentation here as well.

- Database design
- Query Optimization
- Fewer HTTP Requests
- "Expires" Header
- Gzip your output
- CSS at the Top
- Scripts at the Bottom
- Avoid CSS Expressions
- External JavaScript and CSS
- Reduce DNS Lookups
- Minify JavaScript
- Avoid Redirects
- Duplicate Scripts
- Configure ETags
- General code optimization tips

Database design
- Every book will tell you to create the database as Normalized
- Now the brief idea of Normalization is
  - First normal form (1NF)
    - Eliminate Data redundancy
  - Second normal form (2NF)
    - Remove subsets of data that apply to multiple rows of a table and place them in separate tables.
    - Create relationships between these new tables and their predecessors through the use of primary and foreign keys.
  - Third normal form (3NF)
    - Remove columns that are not dependent upon the primary key.
  - Fourth normal form (4NF)
    - A relation is in 4NF if it has no multi-valued dependencies.
- Thumb rule of Normalization and database design or gist of all Normal forms can be
  - Eliminate the data redundancy by having the data organized in such a manner that related data is kept together.

Database Design – Practical approach
- Lets take an example of one of the most happening site on internet – orkut
- If you look at the profile page on Orkut you will see lots of things but lets concentrate on the followings
 - Your profile

- o Your friends (Some 8 or 9)
- Now lets see the what could be the database for such a situation
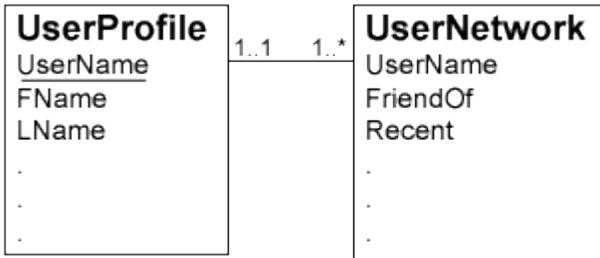
- Now to display the page what you should be doing.
- o Fetch all the data from the UserProfile Table and then
- o Join the table UserProfile with UserNetwork
- o Or Execute 2 queries based on the profile, fetch his friend.
- Now think how costly this join could be when you have millions of users like Orkut and each one of them have thousands of friends.
- You are fetching 8 or 9 records out of millions of them.
- Other solution could be to duplicate the recently used friends data in the UserProfile table itself which will save us querying millions of records.

Query optimization

- As we have seen that design of the database is one of the most important factors in the performance of your web page.
- Better database design also need optimal queries to perform optimally.
- You should also concentrate on designing your query.
- Always think over the query and ask yourself if the same thing can be achieved with some alternate queries and in more efficient manner.
- As an example
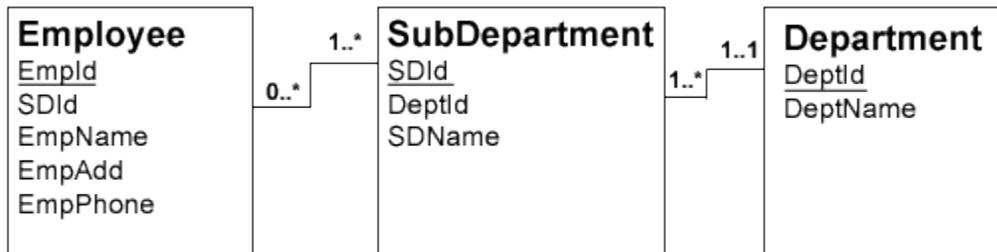
Employee has one or more sub departments
Every Sub-department may or may not have Employee

Sub-Department has one and only one Department
Department have one or more than one department

- Let's take a very common example with the above database,
- o To know all the Employees in a particular department.
- One of the solution could be joining the tables
  Code: SQL

```sql
SELECT e.EmpId, e.SDId, e.EmpName, e.EmpAdd, e.EmpPhone
FROM Employee e, SubDepartment sd, Department d
WHERE e.SDId = sd.SDId
AND sd.DeptId = d.DeptId
AND d.DeptId = 'MyDepartment'
```

- The above solution of joining 3 tables to give output for the MyDepartment as the id of the department is a very common but is a very expensive one.
- Any organization should have "employee to department" ratio high and if we can avoid joining an Employee table to Department table.
- If we analyze the above solution then we are getting the output for only one Department i.e. MyDepartment. So our aim should be to find the Sub-Department's in our MyDepartment.
  Code: SQL

```sql
SELECT SDId FROM SubDepartment WHERE DeptId = 'MyDepartment'
```

- Now all we need is employees in the above Sub-Departments.
  Code: SQL

```sql
SELECT e.EmpId, e.SDId, e.EmpName, e.EmpAdd, e.EmpPhone FROM Employee e WHERE
e.SDId IN (SELECT SDId FROM SubDepartment WHERE DeptId = 'MyDepartment')
```

Advantages of query optimization.

- With the above solution we have avoided joins from the 3 tables.
- Avoided querying Department tables as well.

Fewer HTTP Requests

- When any page is fetched from the server it makes an HTTP request to the server.
- Most of this time is tied up in downloading all the components in the page: images, stylesheets, scripts, Flash, etc.
- Reducing the number of components in turn reduces the number of HTTP requests required to render the page.
- Combined files are a way to reduce the number of HTTP requests by combining all scripts into a single script, and similarly combining all stylesheets into a single stylesheet.
- Majority of web developers' think differently and will not agree on this.

Fewer HTTP Requests – Case Study

- Say I have a program where I need to be doing some disk operations.
- Disk operations are one of the slowest operations and very few will disagree on this.
- Now say we take 2 test cases
- 10 small File I/O's by opening reading and then closing it.

- o   1 time File I/O where the content of the file is loaded in memory.
- Analogy is Web server needs to be doing lots of small Disk operation is to one disk operation.

"Expires" Header

- Expires header makes your web components like scripts, stylesheets, images, and Flash cacheable
- A first-time visitor to your page may have to make several HTTP requests, avoids unnecessary HTTP requests on subsequent page views.
- Expires headers are most often used with images, but they should be used on all components including scripts, stylesheets, and Flash components.
- The major disadvantage of using it is you have to change the component's filename whenever the component changes and the best way of doing is to make the name based on version number of the file or the date on which it was last modified to be added with the file name.
- Using Expires header has an advantage if you have returning users but for the new visitors you have no added advantage.

Gzip your output

- Use of latest browsers helps you send the page content in the zipped format.
- Browser sees the headers and unzips the content and then render it to the screen.
- Gzip is supported by almost all browsers and web servers are also responding to the same and are going hand in hand.
- Gzipping as many file types as possible is an easy way to reduce page weight and accelerate the user experience.
- The disadvantage of using such things is you may not be able to see the web page correctly if you are using a very old browser.

CSS at the Top

- Used correctly most of the time.
- Putting CSS in the HEAD element allows progressively display of web page and the web page itself can behave as a progress indicator.
- Also the layout of the page can change dramatically with the final CSS and so many browser do not render the pages progressively if CSS is at the bottom.

Scripts at the Bottom

- Used incorrectly most of the time.
- Moving scripts as low in the page as possible means there's more content above the script that is rendered sooner.
- The disadvantage is its not always possible to b able to push the JavaScript down and you may have some content in the JavaScript which updates the page (like document.write) and so it should be at the top of the page

- The DEFER attribute specify that though the JavaScript is at the top of the page it can continue rendering and defer the load of the java script file but unfortunately, all browser doesn't support the DEFER attribute.

Avoid CSS Expressions

- Internet explorer supports expressions in CSS which based on some JavaScript expression you can set some values.
- Example
  Code:

```
background-color : expression((new Date()).getHours()%2 ? "#0000FF" :
"#FF0000");
```

- The problem with such expressions is that the expression may just execute the expression infinite number of times when you move your mouse or resize the window.
- Other disadvantage is you cannot use the CSS caching (due to every time changing nature of CSS) which can help you save some HTTP requests on your web server.

External JavaScript and CSS

- One of the most debatable topic is
- Should they go for external to the page
- Makes more HTTP requests.
- Allows caching
- Should be inline with the page
- Less HTTP requests
- My opinion is if we have the option of caching we should go for it
- because it reduces the HTTP requests for returning user.
- The size of the HTML document is reduced without increasing the number of HTTP requests.

Reduce DNS Lookups

- Say we have 10 image servers for the domain go4expert.com as img[0-9].go4expert.com
- Now we have 2 options
- Fetch the images using the host name as img1.go4expert.com
- Fetch the images using the ip address of the server.
- Its obvious to use the ip because it does not need to be resolving the host name saving you some milli seconds.
- The disadvantage of using such method is you should be having static ip address which should not change frequently.

Minify JavaScript

- Its not minimize the use of JavaScript but its to minify the JavaScript.

- Minification is the practice of removing unnecessary characters from code to reduce its size thereby improving load times.
- When code is minified all comments are removed, as well as unneeded white space characters (space, newline, and tab).
- In case of JavaScript, this improves response time performance because the size of the downloaded file is reduced.
- JSMin is one of the open source software available.
- http://www.crockford.com/javascript/jsmin.html
- JSMin is a filter that omits or modifies some characters. This does not change the behavior of the program that it is minifying. The result may be harder to debug. It will definitely be harder to read.
- JSMin removes comments and white space from the code.
- Obfuscation is an alternative optimization method that can be applied to source code which converts function and variable names into smaller strings making the code more compact as well as harder to read.
- The most famous example is iGoogle:-
- Look at the source code of iGoogle page and you will see that lots of JavaScript but you are unable to read or debug it and the reason behind it is that its reduced and is uncommented.

Avoid Redirects

- When you link to your internal pages like
- http://www.go4expert.com/forums
- It actually redirects you to
- http://www.go4expert.com/
- **Note** :- See the trailing slash
- Avoid HTML code linking to the site without the trailing slash to avoid un-necessary redirects on your server.

Duplicate Scripts

- Manage your script module such that you do not include the same JavaScript file into the web page more than once.
- Some browsers make duplicate requests to the server and thus you make an additional HTTP request.

Configure ETags

- ETags helps understand the browser to understand if the cache version of the file is same on the server or modified.
- ETags are created based on file information and server.
- ETags won't match when a browser gets the original component from one server and later tries to validate that component on a different server, a situation that is all too common on Web sites that use a cluster of servers to handle requests.
- If you're not taking advantage of the flexible validation model that ETags provide, it's better to just remove the ETag altogether. The Last-Modified header validates based on the component's timestamp. And removing the ETag reduces the size of the HTTP headers in both the response and subsequent requests.

General code optimization tips

- Avoid using regular expression where possible.
- Avoid declaring variable inside the loops.
- Always go through the documentation of the API's for optimal use of them.
- Some languages allow you can do away with variable declaration. Try avoiding use of such automatic variable declaration system.
- PHP allows you to have the header files inclusion when you need. Do that. Do not go for including all the header files at the beginning.
- Remember OO programming is for better maintainability and not for optimal performance but the performance lost is worth the gain.

Source: http://www.go4expert.com/articles/speed-web-pages-t6814/