



# Digital Wealth

blog.hardeep.name

Computers & Programming

Anniversary Edition

July 2009

Extra: Database  
Performance Tuning

Wordpress How Tos

TEST PLANNING  
LINUX SCRIPTING

HARDEEP SINGH

## **Foreword**

What appears below is a collection of posts from my blog Digital Wealth <http://blog.Hardeep.name> arranged in order for easy following rather than by date. Similar topics have been clubbed together based on the subject, and from basic to more advance.

This eMagazine is being released to commemorate the first anniversary of the blog (16<sup>th</sup> July 2009). While this eMagazine is a snapshot of the blog at one particular time, the blog itself is dynamic: the blog posts change with time, based on comments from readers and my own learnings. Hence, if you like a particular topic, please visit the blog, read the latest – and leave your questions / feedback.

It's not easily possible to reproduce all the hyperlinks in the print version. However, against each post a URL to that post is mentioned. If needed, please enter the URL, read the post online with all links intact.

## **TABLE OF CONTENTS**

<b>ORACLE DATABASE &amp; SQL</b>	<b>1</b>
<b>ONLINE SECURITY</b>	<b>8</b>
<b>LINUX SCRIPTING</b>	<b>18</b>
<b>WINDOWS</b>	<b>21</b>
<b>MATHEMATICS</b>	<b>26</b>
<b>WORDPRESS BLOGGING</b>	<b>29</b>
<b>MANAGEMENT</b>	<b>32</b>
<b>MISCELLANEOUS</b>	<b>36</b>

# ORACLE DATABASE & SQL

## Using SQL potential

By [Hardeep](http://blog.hardeep.name/computer/20080731/using-sql-potential/) • July 31, 2008 <http://blog.hardeep.name/computer/20080731/using-sql-potential/>

I like to use the database to its full potential. For example, suppose someone has a list of vouchers and needs to find the vouchers that were paid later than the due date. One way to do this might be to read the vouchers one by one from the database, compare the due date with the payment date and determine the results. The other, recommended method will be to add the required criteria to the query itself so that only the exact result is obtained. With the second method, only 5% or 10% of the vouchers will need to be transferred from the database to the application while in the first method, all vouchers will need to be transferred.

In other words, the exact business requirements should determine the query. While you are at it, you should also keep in mind the indexes. Queries should always be written to minimise Disk I/O and transfers between the DB and the Application (server).

The database itself is quite powerful (esp Oracle) and I feel that its potential is always under-utilised. Let me show through an example.

I once had a requirement that there is a table having first, middle and last names of employees and the email ID. Something like this, ignoring the datatypes - assume all are VARCHAR2:

```
create table userlist(fname,mname,lname,emailid);
```

Each employee has middle name blank. Its possible that multiple employees have identical fname, lname with each other. For example, there can be two people having name 'Hardeep Singh'. In this case, if the emailid of the two employees is same that means they are the same person having multiple rows, else they are different persons having the same name.

For example:

1. Hardeep Singh alpha@gmail.com
2. Hardeep Singh beta@gmail.com
3. Hardeep Singh beta@gmail.com
4. Satinder Singh gamma@gmail.com
5. Satinder Singh gamma@gmail.com
6. Gorakh Nath gn@gmail.com

In this case, 2 & 3 are the same person and 4 & 5 are also the same person. 1 & 2 are two different people.

Now the requirement is that we have to modify the middle name by adding a number such that every different person has a unique name. In the example above, the names should be:

1. Hardeep Singh '1' alpha@gmail.com
2. Hardeep Singh ' ' beta@gmail.com
3. Hardeep Singh ' ' beta@gmail.com
4. Satinder Singh ' ' gamma@gmail.com
5. Satinder Singh ' ' gamma@gmail.com
6. Gorakh Nath ' ' gn@gmail.com

Now we know that '1' is different from '2' and '3' because he has a different middle name.

The middle name to be added is given at the end of the name, in quotes. Gorakh Nath does not get any middle name since his name is unique. Any Tom, Dick or Harry would do this requirement in the following way: Read all the details one by one, look for people having the same name, then check the emailID then issue an UPDATE like this:

```
UPDATE userlist SET mname='1' where emailID='alpha@gmail.com';
```

Such UPDATES would need to be issued one for each person. However, this can be done through just a single UPDATE statement, without reading the list of employees at all. Here is the query:

```
update userlist a
set mname=(select x from (select rownum x,emailid,fname,
                           lname
                        from userlist xa
                        where exists
                        (select 1
                         from userlist xb
                         where xa.lname=xb.lname and
                              xa.mname=xb.mname and
                              xa.fname=xb.fname and
                              xa.emailid<>xb.emailid))
            ord
        where ord.emailid=a.emailid and
              ord.fname=a.fname and
              ord.lname=a.lname)
where exists(select 1
            from userlist b
            where a.lname=b.lname and
                  a.mname=b.mname and
                  a.fname=b.fname and
                  a.emailid<>b.emailid);
```

I guess an explanation is owed as to how it works. To my knowledge this query would work only in Oracle - but there would be ways to make it work in other Databases as well.

'rownum' returns the number of that particular row in the result set. The 'exists' clause at the end makes sure only people with same names are processed ('gn@gmail.com' is ignored). The part:

```
(select x from (select rownum x,emailid,fname,lname
from userlist xa
where exists
(select 1
from userlist xb
where xa.lname=xb.lname and
      xa.mname=xb.mname and
      xa.fname=xb.fname and
      xa.emailid<>xb.emailid))
ord
```

creates a temporary view having the number, the email ID and the firstname. In the given scenario the result from this will be something like:

1. 1, alpha@gmail.com, Hardeep, Singh
2. This row will be absent because of the xa.emailid<>xb.emailid clause
3. This row will be absent because of the xa.emailid<>xb.emailid clause
4. This row will be absent because of the xa.emailid<>xb.emailid clause
5. This row will be absent because of the xa.emailid<>xb.emailid clause

6. This row wont even be considered, as I explained above

Had there been yet another 'Hardeep Singh' with a different email ID, he would have got a middle name of '2'. Now the last step is to copy over the numbers based on the first and last names only - that part is pretty simple. Please post any questions in the comments area.

## Performance tuning SQL queries in Oracle

By [Hardeep](#) • February 17, 2009 <http://blog.hardeep.name/computer/20090217/sql-tuning/>

Visit <http://blog.hardeep.name/computer/20090217/sql-tuning/>

## Generating sequential numbers in a database

By [Hardeep](#) • February 18, 2009 <http://blog.hardeep.name/computer/20090218/gen-num-sequence/>

You are creating an application that allows organisations to manage employees. One of the tasks that it has to do is generate an employee ID when a new employee is being entered. One way of doing this is through this query:

```
SELECT max(empl_id)+1 FROM employee;
```



Photo by James Cridland <http://www.flickr.com/photos/jamescridland/2272701122/>

However, this query presents a problem in a multiuser environment: if more than one user is entering employee details at the same time, they will both get the same empl\_id. To tide over this problem, one way to go is to look at the auto numbering solution provided by the database - however I personally find that solution limiting and have never used it.

The other approach is to create a single row table for global settings (in all probability your application will already have this) and maintain a field in that table as the last number used. Thereafter, the code can be written as below:

```
UPDATE settings_tbl SET lastnumber=lastnumber+1;
```

```
SELECT lastnumber FROM settings_tbl;
```

Remember that the order of the queries is important in a multiuser environment. Placing SELECT before the UPDATE can cause problems (locking has to happen first).

This piece of code should be executed at the time of saving the employee and not when a request for the blank employee form is generated. This is necessary so that one user of the application doesn't have to wait for another to be finished. Note that the UPDATE lock is released only when you do the COMMIT or ROLLBACK.

Another way of doing the same thing in Oracle, one that I prefer myself and have used in a number of tight situations is the FOR UPDATE clause. This one allows you to do the SELECT first:

```
SELECT lastnumber FROM settings_tbl FOR UPDATE;
```

```
UPDATE settings_tbl SET lastnumber=lastnumber+1;
```

# When NOT to normalise the database

By [Hardeep](#) • March 17, 2009 <http://blog.hardeep.name/computer/20090317/db-not-normalise/>



Photo by Tim Morgan <http://www.flickr.com/photos/timothymorgan/75593157/>

When talking of [Database Normalisation](#), textbooks often talk of [BCNF](#), [fifth](#) and higher normal forms. However, in practice (in large software/ERPs) I have rarely noticed normalisation beyond Third Normal form. In fact, there is a certain degree of redundancy that is desirable.

While doing database design, I believe there are two critical aspects that should be kept in mind but I see ignored in a lot of common software.

The first is the time aspect of data. First - an example from finance. Consider a company having multicurrency invoicing. The tables can be designed as:

INVOICE: InvoiceID, ..., Currency, BaseCurrency, TransactionDate, ...  
CONVERSIONS: FromCurrency, ToCurrency, EffectiveDate, RateMultiplier

This is a design having no redundancy. On the basis of the three fields in the INVOICE relation, we can always find out the latest row from the CONVERSIONS table having EffectiveDate less than TransactionDate. Hence we can determine the RateMultiplier.

Consider another design:

INVOICE: InvoiceID, ..., Currency, BaseCurrency, TransactionDate, **RateMultiplier**, ...  
CONVERSIONS: FromCurrency, ToCurrency, EffectiveDate, RateMultiplier

**Here, the system determines the value of the RateMultiplier at the time of invoice creation and records it permanently within the INVOICE table itself.** To me this would be more mature design. Why? Because a lot of data in the INVOICE table would actually depend on the RateMultiplier: for example the VAT details. If on 1-JAN-2009 we know that the exchange rate is 1.1. However, on 3-JAN-2009 we come to know that the rate was incorrectly recorded. Someone changes the CONVERSIONS table to reflect the new exchange rate, of 1.2. All the details in the INVOICE table for the invoices created between 1-JAN and 3-JAN become inconsistent since the BaseCurrency is now inconsistent with the RateMultiplier.

Now consider an example from HR appraisal systems. A table stores what stage an appraisal process is at for a particular employee. This is then used to decide what access he has.

STAGE\_CURRENT: EmpID, Stage

Note that this has no Date, or Year field. An employee is trying to see records for the previous year appraisals, yet is unable to see some of the data, because current appraisal process is still in initial stage.

The next problem is that of storage of “under calculation” fields. For example, consider the training department maintains the scores of each student trained. The test administered is of 100 marks, but has a weightage 40. Proposed design:

SCORES: CandidateID, TestID, Score, Flag

At the time of recording, the `Flag` is set to `N`. Thereafter a process runs that multiplies the score by 0.4 and sets the `Flag` to `Y`.

In my opinion a better design would be to retain both the scores even though the pre-weightage score is not relevant to the business process, because a process can also terminate in between due to erroneous data being supplied. Hence if the process ends after setting the flag to `Y`, and before changing the score; or in reverse order: after changing the score and before setting the flag then we end up with inconsistent data. Improved design:

Scores: CandidateID, TestID, Score, WeightedScore

At the time of recording, `Score` is entered and `WeightedScore` is set to zero. Thereafter a process runs that multiplies the `Score` by 0.4 and stores the value in `WeightedScore`.

The central idea is to retain all information permanently so that even if the process fails, we know what data existed.

## Oracle deadlocks: the what and the how

By [Hardeep](http://blog.hardeep.name/computer/programming/20090622/oracle-deadlocks/) • June 22, 2009 <http://blog.hardeep.name/computer/programming/20090622/oracle-deadlocks/>

Everyone knows what a **deadlock** is: a situation in which two or more competing processes are waiting for the other to finish, and thus neither ever does. The purpose of this post is to help people understanding the deadlock a little better with a view to enable them to fix the problem when they find one.

Assume that there are two processes running, A & B and that they require a (shared) file and a printer to do their work. Process A locks up the printer, and Process B locks up the file for its own use. Now, none of the processes can complete because they do not have all the resources needed for their completion, and neither will they release the resource they have: they will keep on waiting for the second resource.

**Let us create a deadlock now**, using Oracle database and SQL Plus client.

We opened two sessions, and executed “`set autocommit off`” as the first statement.

Now in the first session we executed:

```
UPDATE ps_voucher SET grp_ap_id='A' WHERE voucher_id='00692096' AND invoice_dt='2-JAN-2002';
```

second session:

```
UPDATE ps_voucher SET grp_ap_id='A' WHERE voucher_id='00692096' AND invoice_dt='13-MAR-2007';
```

back to the first:

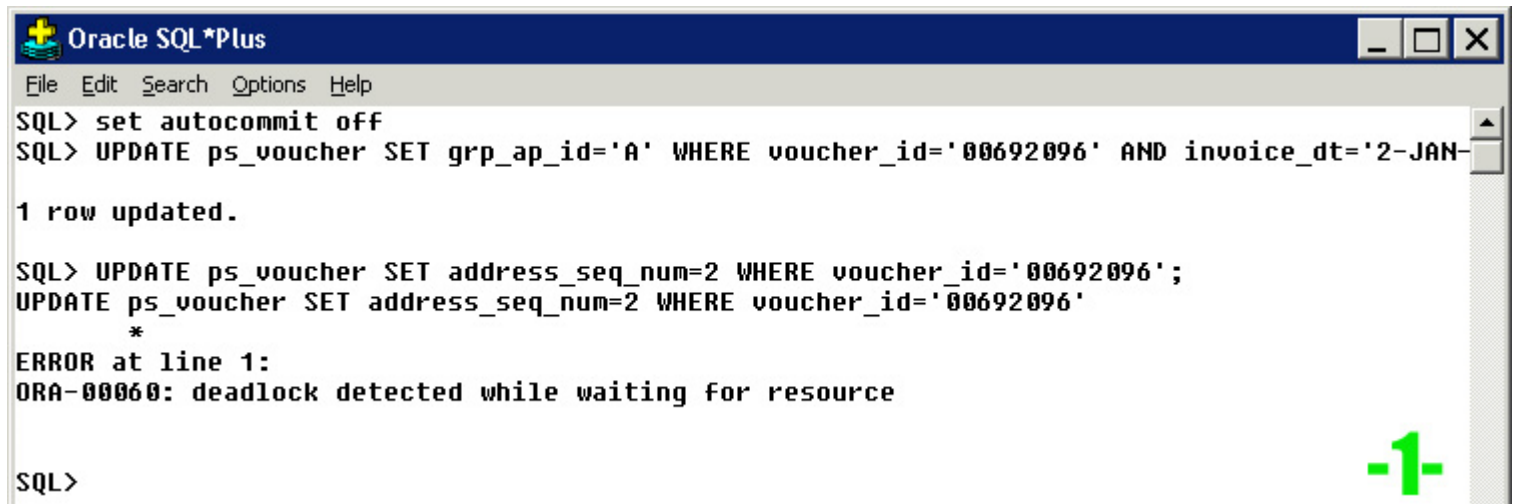
```
UPDATE ps_voucher SET address_seq_num=2 WHERE voucher_id='00692096';
```

and then the second:

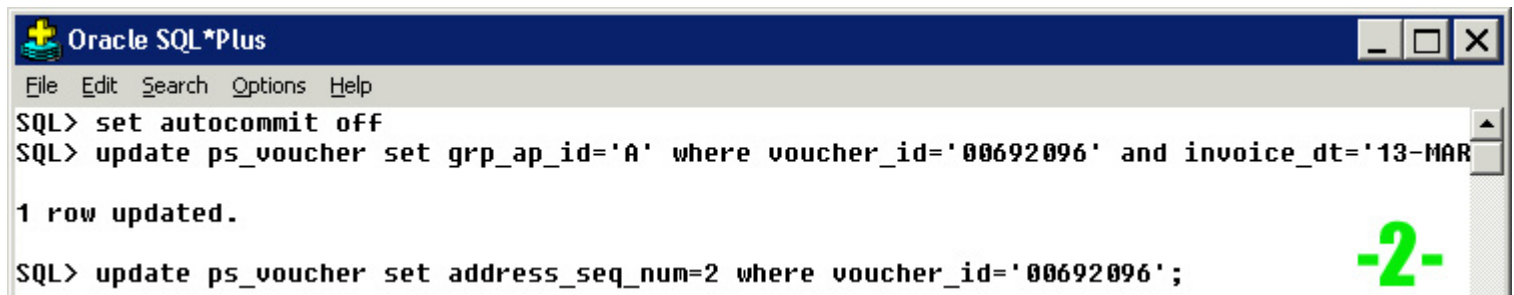


```
UPDATE ps_voucher SET address_seq_num=2 WHERE voucher_id='00692096'
```

**BAM!** Deadlock. See screenshots:

A screenshot of the Oracle SQL\*Plus command-line interface. The window title is "Oracle SQL\*Plus". The menu bar includes "File", "Edit", "Search", "Options", and "Help". The command prompt shows the following sequence of commands and responses:  
SQL> set autocommit off  
SQL> UPDATE ps\_voucher SET grp\_ap\_id='A' WHERE voucher\_id='00692096' AND invoice\_dt='2-JAN-  
1 row updated.  
  
SQL> UPDATE ps\_voucher SET address\_seq\_num=2 WHERE voucher\_id='00692096';  
UPDATE ps\_voucher SET address\_seq\_num=2 WHERE voucher\_id='00692096'  
\*  
ERROR at line 1:  
ORA-00060: deadlock detected while waiting for resource  
  
SQL>  
A large green "-1-" is overlaid on the right side of the screenshot.

Deadlock - Session I

A screenshot of the Oracle SQL\*Plus command-line interface. The window title is "Oracle SQL\*Plus". The menu bar includes "File", "Edit", "Search", "Options", and "Help". The command prompt shows the following sequence of commands and responses:  
SQL> set autocommit off  
SQL> update ps\_voucher set grp\_ap\_id='A' where voucher\_id='00692096' and invoice\_dt='13-MAR  
1 row updated.  
  
SQL> update ps\_voucher set address\_seq\_num=2 where voucher\_id='00692096';  
A large green "-2-" is overlaid on the right side of the screenshot.

Deadlock - Session II

**What went wrong?** There existed two vouchers in the system, with the same `VOUCHER_ID` but with different `INVOICE_DT`s (invoice dates). Each process first locked up one of those vouchers, and then - as the second `UPDATE` - tried to update both. **(On the database side, a process gets a lock on a specific row when it `UPDATES` that row, and the lock is released when the process `COMMITs` or `ROLLBACKs`.)**

Yes, the programmer could have been smarter and written better code: if he had put the `INVOICE_DT` clause in the second statement also we would have been fine. However, in practice, with huge systems having tons of code - programmer will sometimes make mistakes. Even if they do not, deadlocks will occur: not all deadlocks are caused by SQL issues.

From a system design perspective, **what can be done to prevent deadlocks?** One way is for the execution of each process to have a unique ID - let's call it process instance (PI). So if a process ABC is run once, it will have a PI of 1222 and when it's run next it will have a PI of 1224. If, after this process PQR is run, it will have a PI of 1223. Before changing any transactions, the process can update its own PI on the transactions that qualify:

```
UPDATE ps_voucher  
SET pi=1223  
WHERE <process specific selection criteria>  
AND pi=0;
```

```
COMMIT;
```



The commit here is important - only then will other processes be able to see the 'locking'.

Thereafter the normal processing SQLs can be changed as below:

```
UPDATE ps_voucher
SET grp_ap_id='1'
WHERE <process specific criteria>
AND pi=1223;
```

At the end, set the transactions back to 'open for processing' by setting PI to zero:

```
UPDATE ps_voucher
SET pi=0
WHERE pi=1223;
```

If there are other ways to achieve this, please let me know by posting comments.

The DBA is usually able to specify the SQL queries involved in a deadlock. Many times one process is UPDATIng the rows that the other is DELETIing.

## ONLINE SECURITY

### SFTP>FTP+GPG

By [Hardeep](#) • August 4, 2008 <http://blog.hardeep.name/computer/20080804/sftp-ftp-gpg/>

Recently, we needed to setup secure file transfer with a third party that did not support [SFTP](#). Hence we decided to use FTP+GPG - which includes transfer of [GPG](#) encrypted files with FTP. For more details on GPG, visit the homepage <http://www.gnupg.org/>. It can also operate in 'portable' mode - from a USB key.

We did not realise that this combination is still significantly less secure than SFTP based transfer. The threats it does not cater to, which SFTP can combat are:

- Someone can come to know about the username, password used in the connection since this is still being transferred un-encrypted
- Using this, someone can load their own files into the server which can be designed to either infect the system, or upload malicious information

Even if GPG signatures are used (as we did after coming to know about these shortcomings), there is still the risk of someone replaying the same files that we received. Even this can cause problems in certain situations. Hence, we had to resort to fixing the source IP to combat the problems. Due to using signatures and IP fixing, it all turned out to be more cumbersome, and still less secure.

The best way currently to establish automated file transfers is to do a password-less SFTP. It does not require any password transfer over the network at all, keeps everything encrypted, and: a disgruntled employee who happens to remember a password (in case of password based SFTP) cannot disrupt your services.

This document contains instructions on how to set this up: [Password SFTP - How to setup](#). There is also an [online](#) version of the same document. If you have any questions, please post as comments.

# [Automate encryption with GPG](#)

By [Hardeep](#) • September 4, 2008 <http://blog.hardeep.name/computer/20080904/auto-gpg/>

This blogpost requires some familiarity with [GPG](#).

Today I want to share the scripts for running [GPG](#) in the batch (unattended) mode. You can have a password on the keys if you want, but since this is the automated mode, you may want to use keys without a password. Irrespective of whether or not you use a password - use a separate set of keys that you will not use for anything that not batch processed. The scripts are primarily for Linux. If you need them for Windows, please [read this](#) and post any problems you face in comments section, and I will help.

The script for encryption is here:

```
#!/bin/sh
GNUPGHOME='/apps/gpg/'
export GNUPGHOME
gpg --batch -r <recieipient> --output $2 --passphrase-fd 3 --sign --encrypt $1
3</apps/gpg/passph
```

The first line shows the location of the shell - please change it according to where the shell is on your system. The second line has the location of the folder where the keys are located - the pubring.gpg and secring.gpg. In the last line, replace <recieipient> with the name on the recieipient key. The /apps/gpg/passph points to the location of the file containing the passphrase. The script will both sign and encrypt the file - change this to suit your needs. The script expects the name of the input file and the name of the output files as parameters in that order.

The script for decryption is here:

```
#!/bin/sh
GNUPGHOME='/apps/gpg/'
export GNUPGHOME
gpg --batch --passphrase-fd 3 --status-fd 1 --decrypt $1 3</apps/gpg/passph | grep '\[GNUPG:\]
GOODSIG'
if [ $? -eq 0 ]; then
    gpg -batch -passphrase-fd 3 -output $2 -decrypt $1 3</apps/gpg/passph
fi
```

The first three lines are similar. Line 4 just checks if the file has a valid signature. If you want to skip this step remove lines 4, 5 and 7. No effort is made to see who signed the file. In my scenario, I could control the people who could sign by having only those public keys in the repository, and the repository could only be written to by 'root'.

Please post comments in case of questions, concerns.

## 7 Responses to “ Automate encryption with GPG ”

1.

User [Feb 15th, 2009 at 5:58 pm](#)

Hi,

I am working on decryting a pgp file using GnuPG.I want to do the same in a .NET C# Console Application. I am facing some problem,when i try to decrypt it prompts for the PassPhrase.I tried to pass the Passphrase from the application but its not working.I want to decrypt the file,stream the dataout.My code is below.If you can kindly help me on this.Here i manually put the passphrase & then i get the result in the Temp string variable.

---

```
using System;
using System.Collections.Generic;
using System.Text;
using System.Diagnostics;
using System.IO;
using System.Threading; // for Thread class

namespace ConsoleApplication3
{
    class Program
    {
        static void Main(string[] args)
        {

            string passphrase = "$Trans@RtA09";

            Process myProcess = new Process();
            StreamWriter sw;
            StreamReader sr;
            StreamReader err;

            ProcessStartInfo myProcessStartInfo = new ProcessStartInfo(@"C:\gnupg\gpg.exe");
            myProcessStartInfo.Arguments = "-decrypt C:/Transfer/test.gpg";
            myProcessStartInfo.RedirectStandardError = true;
            myProcessStartInfo.RedirectStandardInput = true;
            myProcessStartInfo.RedirectStandardOutput = true;
            myProcessStartInfo.UseShellExecute = false;

            myProcess.StartInfo = myProcessStartInfo;

            myProcess.Start();
            sw = myProcess.StandardInput;
            sr = myProcess.StandardOutput;
            err = myProcess.StandardError;

            sw.AutoFlush = true;
            if (passphrase != null && passphrase != "")
            {
                sw.WriteLine(passphrase);
            }

            sw.Close();
            String Temp = sr.ReadToEnd();
            Temp += err.ReadToEnd();
        }
    }
}
```

---

Hi,Thanks for replying.

I tried,but no Luck  
below is the update code which i am using now.

---

```
using System;
using System.Collections.Generic;
using System.Text;
using System.Diagnostics;
using System.IO;
using System.Threading; // for Thread class

namespace ConsoleApplication3
{
    class Program
    {
        static void Main(string[] args)
        {

            // string passphrase = "$Trans@RtA09";

            Process myProcess = new Process();
            StreamReader sr;
            StreamReader err;

            ProcessStartInfo myProcessStartInfo = new ProcessStartInfo(@"C:\gnupg\gpg.exe");
            // myProcessStartInfo.Arguments = "--decrypt C:/Transfer/test.gpg";
            myProcessStartInfo.Arguments = "-batch -passphrase-fd 1 --decrypt C:/Transfer/test.gpg < C:/Transfer/passph.txt";
            myProcessStartInfo.RedirectStandardError = true;
            //myProcessStartInfo.RedirectStandardInput = true;
            myProcessStartInfo.RedirectStandardOutput = true;
            myProcessStartInfo.UseShellExecute = false;

            myProcess.StartInfo = myProcessStartInfo;

            myProcess.Start();
            sr = myProcess.StandardOutput;
            err = myProcess.StandardError;

            String Temp = sr.ReadToEnd();
            Temp += err.ReadToEnd();

        }

    }
}
```

---

If you know of any other approach of doing this in windows application environment?I mean decrypting the

pgp file,by passing passPhrase from application,so it doesn't prompt. and finally reading the output of the decrypting file.

Thanks Again.

3.

User [Feb 16th, 2009 at 6:41 pm](#)

Hi Hardeep,

Do i need to write "-batch" in the command?

4.

[Hardeep Singh Feb 17th, 2009 at 1:04 pm](#)

Some of the options are appearing in the blog with a single dash - however, everywhere there should be a double dash. So there should be two dashes before batch, and two dashes before password-fd.

Can you redirect the output from the gpg command and show what it says?

Otherwise, try the batch process approach that I suggested.

5.

[admin Feb 27th, 2009 at 4:14 pm](#)

@User

What I prefer to do, instead of calling GPG directly from the code is to create a batch file/shell script, and call the script from code. That is the best approach however, try changing this:

```
myProcessStartInfo.Arguments = "-decrypt C:/Transfer/test.gpg";
```

to:

```
myProcessStartInfo.Arguments = "-batch --decrypt C:/Transfer/test.gpg";
```

Also, you will need to keep the passphrase blank with this approach. If it doesnt work you will need to do this:

```
myProcessStartInfo.Arguments = "-batch -passphrase-fd 1 --decrypt C:/Transfer/test.gpg  
<c:/Transfer/passph.txt";
```

Here the file c:/Transfer/passph.txt should contain the passphrase. Let me know what happens. YMMV (because you are on Windows).

6.

Javeed [Mar 20th, 2009 at 5:27 am](#)

try this...

```
try
{
System.Diagnostics.ProcessStartInfo psi =
new System.Diagnostics.ProcessStartInfo("cmd.exe");
psi.CreateNoWindow = true;
psi.UseShellExecute = false;
psi.RedirectStandardInput = true;
psi.RedirectStandardOutput = true;
psi.RedirectStandardError = true;
psi.WorkingDirectory = "C:\\Program Files\\GNU\\GnuPG";

System.Diagnostics.Process process = System.Diagnostics.Process.Start(psi);
// actually NEED to set this as a local string variable
// and pass it - bombs otherwise!
string sCommandLine = "echo " + _passphrase +
"| gpg.exe --passphrase-fd 0 -o \"" +
outputFileNameFullPath + "\" --decrypt \"" +
@"C:\Program Files\GNU\GnuPG\test.txt.gpg" + "\"";
process.StandardInput.WriteLine(sCommandLine);
process.StandardInput.Flush();
process.StandardInput.Close();
process.WaitForExit();
process.Close();
}
```

7.

J. Afgann [May 18th, 2009 at 9:35 pm](#)

Really insightful posts here. Nice info on GPG

## **DNS poisoning - in plain English :-)**

By [Hardeep, via email](#) • November 26, 2008 [http://blog.hardeep.name/computer/security/20081126/dns\\_poisoning/](http://blog.hardeep.name/computer/security/20081126/dns_poisoning/)

Computers do not have names, only numbers. For example, [Yahoo!](#) is 68.180.206.184. When you ask your browser to be taken to Yahoo.com, it looks up this number in something similar to a telephone directory.

A person by the name Dan Kaminsky has found that it's possible for someone (called X henceforth) to modify the directory. So, when you ask for Yahoo, the computer would lookup the wrong number, a number to a computer owned by X. Thereafter, it can record the information you give it, thinking that its Yahoo!. The owner may then misuse this information.

**So what do you do to make sure this doesn't happen?** First, check your DNS server (the machine that tells your computer what number a website has) by going to <http://www.doxpara.com/>. If that says your computer is vulnerable, change your DNS servers. Detailed instructions for different operating systems are [here](#). For Windows, in short you have to change your DNS settings to 208.67.222.222 and 208.67.220.220 by going to Network Connections.

Note: This is not a security update blog, and I do not talk about every vulnerability. However, this one is important and everyone needs to understand it. On the other hand, most forums on the web talk highly techno. Read [this one](#) for more information, only if you think you need it!


## **Safe Browsing Guide**

By [Hardeep](#) • January 27, 2009 <http://blog.hardeep.name/computer/20090127/safe-browsing-i/>

When you browse over the Internet, or Chat, or send/receive email - you are not doing that in private. It is important to understand exactly what is private, and what steps you need to follow to maintain the privacy.

When accessing any website such as Yahoo.com, you get connected to the web site's server which provides you the information you seek – search results, or email. This connection is not direct: you are connected through a series of nodes. Each node can view/alter the information that is flowing through it.

A [protocol](#) - 'https' provides privacy to your interaction by adding a '[Secure Socket Layer](#)' on top of the normal HTTP protocol. Enough of jargon, back to English!

So when you use this particular protocol you are secure subject to some caveats. Use of this protocol can be confirmed through the 'https' at the beginning of the URL, and through the 'lock' icon at the bottom right: .

A lot of online websites support HTTPS for logging in. You have to select 'secure' at the login screen where you enter username/password. This means that your password is protected during the communication. However, these sites move back to normal mode after the login: your data (for example the email content) is not protected. Gmail supports secure connection even after login but you have to enable it in the settings – this makes sense and you should do it. However, even after doing this it does not mean that all your content is 'protected' – more on this later.

Please understand that if its emails in question: just your using a secure connection is not enough. The recipient should also use it for the information to remain inaccessible at the nodes.

You might get some 'warnings' while opening secure pages, like below:





The first one means that the certificate could have been signed by anyone, including by someone you do not trust. If all I need is encryption: for example if I am sending email, and not sending any corporate presentations, or password, or credit card numbers – I should be fine. Many websites use such certificates – but if it's a bank website using it – I would not login. To give you an idea of the relative security impact of many of the warnings discussed during this guide: this particular warning has a severity of 40 out of 100. It means even on seeing this warning, I am 60% likely to use the website.

The second one means that the certificate has expired. In this case, I would check what is the expiry date (by clicking on the lock) and if it expired a couple of days back I would allow it. In addition, if its sensitive information, I would not use the site until the issue is fixed. This one has a severity of 30.

The last one says that the website names do not match between the certificate and the actual website you are visiting. This could be a case of [phishing](#), and could be serious. What I do in such a case is I find out what name the certificate is issued to (clicking on the lock icon) and check it against the actual website in the URL. If the website visited is **server.icicibank.com** and the certificate is for **icicibank.com** – I continue to use the site. If the two are very different, I don't use the site. Severity is 80.

Mozilla issues one message for each of the issues as noted above while Internet Explorer (shown above) issues just one message with error icons:

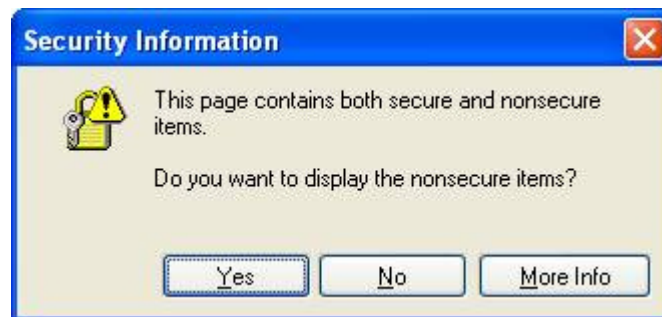


This corresponds to the second warning from Internet Explorer.



This one corresponds to the first warning from Internet Explorer, same actions apply.

Another warning that you might see is this one:



Correspondingly for Mozilla:

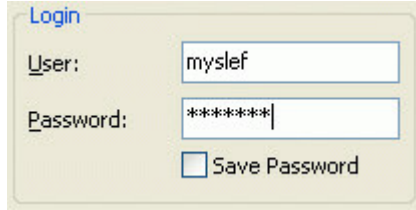


This warning means that there is some content on the page which is not encrypted: this could be images or something else. Severity is 50: on GMail it may mean that the emails have such content so it's ok to continue to do your stuff. However, if on one of the emails you are sending a password – you may want to be careful.

**Disclaimer:** This is just a guide and is not meant to replace professional advise. No measures can guarantee 100% security. There are a lot of threat vectors outside the scope of this tutorial: such as key loggers on your computer. In addition, the severities explained for warnings are just guides and have no scientific basis.

## Remembering your passwords

By [Hardeep](#) • February 21, 2009 <http://blog.hardeep.name/computer/security/20090221/passwords/>



### Remembering passwords

In today's online world we have to remember a lot of passwords. Security advice is that they must all be different (so that if one account is compromised, we don't lose the security of all accounts), and be changed every now and then.

It is sometimes difficult to cope with such requirements, without writing down the passwords - and if you do write them down, what happens if you lose your notebook?

The answer to such problems is a piece of software called 'Password Safe' - it has been developed under the watchful eyes of a well known security expert. It keeps all your passwords encrypted by one single password. So you only need to remember one password from now on. You can keep the software on a USB pen-drive (its even platform independent so you can run it both on Windows and Linux) and take it along with you - no harm if you lose the pen-drive - no one can retrieve your passwords.

Note that there are a lot of different software applications with the same name, and not all of them may be equally secure. Please download only from <http://passwordsafe.sourceforge.net/>.

Note: this is not professional advice, use at your own risk.

## False sense of security

By [Hardeep](#) • June 23, 2009 <http://blog.hardeep.name/computer/security/20090623/false-security/>



Photo by public15 <http://www.flickr.com/photos/publik15/3316679657/>

<http://www.e-signature.com/> provides logo and signature embedded fonts. Per se, there is nothing wrong with the service, and is useful for people who need it.

However, a look at the website throws up the claims below. Each one has a quote from me below it on the truth behind the statement.

Imagine being able to distribute a signed word processing document to your staff or clients without fear of the logo or signature being copied...

**DW:** Even if the font is embedded, the logo/signature can still be copied and used on other documents. While this is not possible as a font letter, it can be done as a bitmap.

No embedding is the option most often selected for signatures as it ensures the highest security - the signature will only appear on computers that have the font installed. If you use signature fonts for signing checks, no embedding is also recommended.

**DW:** Please don't use this for cheques. Its very easy for someone to access your computer and steal the font than you think. The font will be installed on your PC and will remain there, its not something that you can carry on a pen drive and keep in lock and key.

The recipient does not have access to the font to reproduce it in another document nor can they cut-and-paste the on-screen signature or logo.

**DW:** Again wrong. It cannot be cut-and-pasted as a font letter, but can be done as a bitmap.

Keep your font proprietary. No one can copy your font because it will reside in an inaccessible cache on the browser. You can display your text the way you want without fear of anyone copying it.

**DW:** What's an inaccessible cache? No cache is inaccessible - a smart user can always take your font away. Remember: if the browser can download it (and use it to show your content), a user can download it too.

It appears to me that these folks are very much aware that they are making statements that are not 100% truth. A false sense of security is worse than no security.

## LINUX SCRIPTING

### Scripted thumbnail generation: security perspective

By [Hardeep](#) • April 4, 2009 <http://blog.hardeep.name/computer/20090404/image-thumbnail-secure/>



While searching for something on the net, I came across some scripts that generate image thumbnail on the fly.

For example: <http://tech.mikelopez.info/2006/03/02/php-image-resize-script/>.

While using such scripts we should be aware of the security point of view: your site can easily become a proxy for other people or websites.

Another similar script is here: <http://www.findmotive.com/2006/08/29/create-square-image-thumbnails-with-php/>.

The normal way to use these scripts is to upload the script onto a web-server (let us say at <http://domain.com/imgsize.php>). Then, wherever you need to include images on your webpage (for example <http://domain.com/image.jpg>), you need to add something like:

```
<img src='http://domain.com/imgsize.php?img=http://domain.com/image.jpg' alt='test' ... />
```

that is,

```
<img src='<path_to_script>?img=<path_to_image>' ... />
```

The entire point behind using the script rather than directly using the image in the `<img>` tag is to save some bandwidth, and to make the page faster loading. The original image is larger and we need a smaller version on the page, so rather than letting the client browser reduce the size, we reduce it on the server.

However, by doing this, you are lending your domain to those people who want to route illegal traffic. Or, they can use your script to access traffic from behind a proxy that prevents certain images to be accessed. This can either affect your website's credibility or burn out your bandwidth.

Does that mean such scripts should not be used? No, they can be used after incorporating some security into them. There are various ways to do that, and one of the simplest ways of doing so is explained below.

The page where the URLs are being generated (so where the IMG tag is being generated) should 'sign' the URLs like so:

```
$url = '...';  
$sign = md5( 'password' . $url );
```

(replace password with the password that you want to keep)

This signature should be added to the URL for the image:

```
<img src='<path_to_script>?img=<path_to_image>&passph=<the_sign>' ... />
```

The script (imgsize.php) should be modified to recompute the signature and match the recomputed signature to the one that comes from the URL. If they match, provide the access and if not, display an error message.

Please contact me through comments if you need assistance on implementing this approach. This suggestion does not come with any warranty (please use it at your own risk) and it does not provide 100% protection. If someone is able to guess your password, or look at your source code, he can also generate signatures just like you can. However, this does provide a minimal security layer that was missing in the original script: you have locked your house.

## [Searching files within multiple folders](#)

By [Hardeep](#) • September 2, 2008 <http://blog.hardeep.name/computer/20080902/search-folders-content/>

On Linux, the normal way to search for some text within a file is to use [grep](#) (Global Regular Expression Print). However, `grep` has a limitation: it cannot automatically search folders within the current folder. It can only search within files in the current folder. **Today I will show you how to use `grep` to search within all files and folders inside a current folder (recursively).**



**Windows users - despair not.** If you find the standard windows search brain-dead or want to automate the task through scripts, you can also use this script. I have [already explained](#) various ways to run Linux scripts on windows - use the one that suits you.

We will couple `grep` with the [find](#) command, to unleash the power.

Here is what you need to do on Linux (or Cygwin):

```
find . -type f -exec grep -iH 'dedicated' {} \;
```

and use this for UnxUtils on Windows:

```
find . -type f -exec grep -iH "dedicated" "{}\" ;
```

This does a case insensitive search for the word ‘dedicated’ in the current folder and all subfolders under it. Change `-iH` to `-H` for case sensitive search.

You can read the manuals for `find` and `grep` and change the commands to suit your needs - this method provides a lot of flexibility. Post your precise usage in comments, especially for Windows.

## Deleting zero byte files

By [Hardeep](#) • September 11, 2008 <http://blog.hardeep.name/computer/20080911/del-zero-byte-files/>

In the past I have shown ways to [run Linux scripts on windows based system](#). I have also talked about one [use of the find command](#), in conjunction with `grep` command to search for files having a given text content (say a word) in multiple folders.

Today I will show you another use of the `find` command: to automate tasks such as deleting zero byte files. This is a pretty common cleanup task that’s carried out on machines that are involved in EDI file transfers.

Here is the script:

```
find . -size 0 | sed -e 's/^/rm /' | sh
```

It deletes all zero byte files in the current folder, and in the folders below it.

In order to understand this, you may need to read about [pipes](#). The task is carried out in three steps:

1. `find . -size 0` searches for all zero byte files in the given folder and the folders below it and returns the filepaths.
2. `sed -e 's/^/rm /'` turns the list of names into a script - for example if the name is ‘/data/x’ it changes it to ‘rm /data/x’. More on [sed here](#).
3. the last steps simply forwards the script to the Linux shell for execution.

This is also a very flexible script and can be customised to carry out a wide variety of tasks. Please post your variations as comments.

## Moving multiple files

By [Hardeep](#) • October 10, 2008 <http://blog.hardeep.name/computer/20081010/moving-multiple-files/>

One common requirement on \*nix systems is to rename (or move-and-rename) more than one file, matching a wildcard. However, if you try to do this:

```
mv ACK*.xml /code/xml/ACK*.xml.done
```

it won't work for more than one file. By the way, this is one of those odd places where Windows command line works better!

How to work around this? As below:

```
ls ACK*.xml | sed -e 's/./mv & \/code\/xml\/&.done/' | sh
```

I have [already talked](#) about sed, sh and this style of scripting: **it can work also on Windows**. In the target path all slashes need to be escaped like this: \ since its being used inside a sed script. Of course, this was easy because we had to suffix something to the filename - had we needed to add something in the middle, it would have been more difficult (and a bigger sed script would be needed).

Here is a more 'capable' [script](#): it takes two parameters, first one the path to the source directory, and next one the path to the target directory. It moves ACK\*.xml files from source to target/.processed (.processed folder under target) renaming them with the current datetime stamp.

Customise it to suit your needs.

## WINDOWS

### NTLM authentication proxy

By [Hardeep](#) • September 29, 2008 <http://blog.hardeep.name/computer/20080929/ntlm-authentication-proxy/>

What would you do if you had a web proxy that requires [NTLM](#), and the software (for example, HTTrack) that needs to connect to the Internet, doesn't support it? NTLM by the way, is an authentication protocol used by Microsoft.

I had this issue, and downloaded [a proxy server](#) (APS) that can connect to an NTLM proxy. The given software (HTTrack in this case) needs to connect to this new proxy server. It works flawlessly. You will need to download [Python 1.5.2](#) - which is an older version of Python. The proxy server can only work under that particular version of Python. For example if APS is running on the same machine as HTTrack (port 6000) and your real proxy is on machine NTLMMain, port 9000 - enter 'NTLMMain:9000' as configuration for APS, and enter 'localhost:6000' in HTTrack configuration.

### NTFS: Alternate Data Streams

By [Hardeep](#) • October 15, 2008 <http://blog.hardeep.name/computer/20081015/ntfs-ads/>

Not many people working on Windows using the [NTFS file system](#) are aware of a feature called Alternate Data Streams (ADS).



First things first: A file system is that part of an operating system which deals with storage and management of files on the disk. Its responsible for the ability to retrieve data from a file when needed. An operating system can support multiple file systems - for example, Windows XP supports at least FAT and NTFS.

Now we come to ADS. Its a feature of NTFS, and allows data to be stored in “hidden” blocks linked to a file. Normal operations such as reading or writing to the file do not reveal the existence of this data. So it can happen that your hard disk is full, yet the file sizes do not reveal the reason. Here is a small ‘test’ I did in this regards (for people who understand command line - a more ‘Windows’ example is available [here](#)):

```
C:\>copy con hello.txt
This is visible to all.
^Z
        1 file(s) copied.

C:\>dir hello.txt
Volume in drive C has no label.
Volume Serial Number is A8EB-38D4

Directory of C:\

10/15/2008  12:56 PM                25 hello.txt
              1 File(s)                25 bytes
              0 Dir(s)  8,501,428,224 bytes free

C:\>echo This is visible only to some > hello.txt:hiddendata

C:\>type hello.txt
This is visible to all.

C:\>dir hello.txt
Volume in drive C has no label.
Volume Serial Number is A8EB-38D4

Directory of C:\

10/15/2008  12:56 PM                25 hello.txt
              1 File(s)                25 bytes
              0 Dir(s)  8,501,428,224 bytes free

C:\>more < hello.txt:hiddendata
This is visible only to some, and is part of ADS
```

This is what I did: I created a file called hello.txt with some text. Thereafter, I checked the file-size and it was 25 bytes. I then added some text to the file, but under an ADS named ‘hiddendata’. I again checked the file-size: it was still shown as 25 bytes, and the contents did not show the content of the ADS. However, when I directly asked to be shown the content of the ADS, it was shown to me.

Windows may use ADS to store access related information, or anything it wants to. In particular this is used to store “zone” information - when a file is downloaded from the internet and you see a warning when executing the file - its because of this zone information stored in the ADS. To remove ADS from a file, a quick way is to copy it to a FAT device and then copy it back. Since FAT cannot hold ADS, it gets deleted.

Virus writers, however, may use this feature for anything: they can store their own files without anyone noticing. Can you run an executable file directly off an ADS? You bet:

```
C:\>start /b c:/hello.txt:hello.exe

C:\>
Hello, World.
```

Hence it's important to keep an eye on your NTFS drives. One way to do this is to use a tool called [LADS](#). It comes with a very nice [FAQ](#) on ADS. Another tool, one that I haven't tried myself is [LNS](#).

The official word from Microsoft is [here](#), and another interesting PDF is [here](#).

LADS can be run on a partition, say C: as below:

```
lads c: /s
```

If you notice anything strange beyond "Zone.Identifier" or "Thumbs.db:encryptable" - especially a file with 'exe' extension - it can be a cause for concern.

## [Microsoft kills DHTML to take on Google](#)

By [Hardeep](#) • October 30, 2008 <http://blog.hardeep.name/computer/20081030/microsoft-kills-dhtml/>

Back in the yesteryears, Microsoft wanted to build a great browser - hence it ensured it was leading the DHTML feature list. Sites began to be developed specifically for Internet Explorer version 4 or 5. At some point however, it realised it had made the browser too powerful - and Google was taking advantage of that, releasing software as service.

Microsoft obviously hates software as service model since it doesn't tie people down to a particular operating system. People could very much use a Linux based browser (such as Firefox) to access those services, and at the same time share documents with Microsoft Windows based users who could access those documents as well through Internet Explorer.

One of the strategies Microsoft uses to kill competition is to not support their USPs on Windows. Since most users use Windows **today**, that prevents those USP features to gain acceptance and die sooner or later. It understood that it could not do this with DHTML: obviously because on the one hand DHTML was its own creation and on the other it would break backward compatibility with Internet Explorer. Hence, it decided to at least halt the development of DHTML and has [not added any major features to DHTML](#) in the recent past.

Just an opinion.

## [Restoring your system after a crash](#)

By [Hardeep, via email](#) • November 29, 2008 <http://blog.hardeep.name/computer/20081129/restoring-a-crash/>

Ok, so your computer crashed. You know one way out: reformatting the hard disk - but you don't like it. So what can be done? If your data is really important, and you are prepared to pay some hard bucks getting it back, talk to a professional. Period.

Still here, ok - so your data is important, but you want to do what you can, yourself. I will discuss in the blogpost some tools that can help.

One of the best recovery tools, that has helped me as well, is: **Emergency Boot CD** (EBCD). The original site seems to be no longer active, and it hasn't been updated since 2004, but its still a great tool. Look for it [here](#), or [here](#), or [Google search](#) for it.

With this, you can recover deleted files, or those lost by formatting. You can fix master boot record. It will allow you to boot from a specific partition, even if its being shown 'non-system' otherwise. Once you have the system up through EBCD, on XP you can run the 'bootok.exe' command, or 'fixmbr.exe' to fix any issues, its available in /windows/system32.

EBCD includes Windows Password Wizard, which can be used to restore access to your PC when you forgot the password for Windows user account.

The next option I am going to talk about is “Bart’s Preinstalled environment bootable live CD” ([BartPE](#)). However, using this is not just a matter of downloading an ISO and burning a CD, as it was for the EBCD. You will need to create the CD from your Windows XP CD. The instructions are on the [BartPE](#) website. Once you have the CD ready, you can boot off the CD into a live Windows XP environment with a lot of tools available. You can even access NTFS drives. See where you can get with this. Again, once you have the system booted, try running bootok.exe as I suggested above.

Another option, if you cannot/do-not-want-to get the BartPE built, is a [Knoppix Live CD](#). This will boot your computer with Linux OS, which is not very difficult to use given Windows knowledge. However, other than getting your computer running, this will not do much more to fix your problem.

The last recovery tool I am going to talk about is [System rescue disk](#). This one offers the GParted tool, which is a disk partition management tool (similar to FDISK) but can resize partitions without losing data.

**If you are unable to download/lay your hands on one of these CDs, I can have them shipped out to you for a fee (except, of course the BartPE).**

## [Is the Windows registry a good idea?](#)

By [Hardeep](#) • December 7, 2008 <http://blog.hardeep.name/computer/20081207/windows-registry-good/>

Compared to Unix config files, and even to Windows 3.1 ‘ini’ files is the windows registry a good idea? This was the question that presented itself in my mind this weekend.

First, **what is the windows registry?** Windows registry is a [Hierarchical database](#) that stores configuration settings for Windows, and for the software installed on your machine.

Now, coming back to the original question I feel while at the core, **the registry is a good idea** - keeping every setting in a centralised place - it **encourages non-portable software**. If you install, say, CorelDraw at a certain location on your harddisk, and want to move it to a different location, you need to uninstall it and then reinstall again. Ditto for moving from one computer to another. A time-consuming process - especially if you use multiple computers - one at work, one at home, and another at a cafe.

**Why is it that I say the registry, per se, encourages non-portable software?** The reason is, when you had config files, and you moved software, the files moved with it too. With registry this doesn’t happen. Note however, either way its possible to write portable or non-portable software. Just that, when a programmer writes code without portability as a key focus, and uses config files, the end product is more likely to be portable.

**What can be done about it?** When a software program runs, it needs to check if the keys it needs are present or not. If not, it should try to default the parameters, and add them to the registry.

One current approach to writing portable software, is to use config files and provide an update program: if you move the software, you run the update program. This program will detect the file-paths and update those in the config files. The same can be used for registry.

## [Portable software](#)

By [Hardeep](#) • August 26, 2008 <http://blog.hardeep.name/computer/20080826/portable-software/>

[Portable software](#) refers to programs that can be stored onto a media (such as USB drives, CDs, external harddisks etc.) and run directly from there on multiple computers. In other words, **software that you can move from one computer to another without the need to re-install.**

There are a lot of [sites](#) that provide such software for download. Google for “portable software” in general, or, say for “portable firefox” in particular. My favourite portable apps are the GIMP, and portable FileZilla (a GUI FTP tool).

[Cygwin](#) is a Linux like environment for Windows. It can run within Windows and access the filesystem. Its very useful to people like me who need Windows as the main OS, but need to test Linux shell scripts and other utilities. I even have a portable version of Cygwin, created based on [these instructions](#). In fact, I was able to improve the procedure slightly. If someone needs help, please contact me through comments. I want to upload the ISO of the DVD created - if someone can provide the bandwidth and hosting space, I can mail a copy of the DVD to him.

## [Running Linux scripts on Windows](#)

By [Hardeep](#) • August 28, 2008 <http://blog.hardeep.name/computer/20080828/linux-shell-on-windows/>

Running Linux shell scripts on Windows is very useful to me. We may need to tweak some scripts depending on the environment in use.

One of the best ways to run them, is within **Cygwin**. This is something I have already [written about](#).

If Cygwin is too bulky for you, you can **use [UnxUtils](#)**. This is a small set of widely used Linux Utilities, and includes the zsh shell. If you use this you will need to limit the scripts to those that use the commands available within this set. However, it still packs in a lot of paunch.

You might need to ensure the file path syntax expected by the given script matches the environment. Cygwin supports both forward-slash and backward-slash. However, please check the documentation/test to see what works.

Last option, is to run **Linux on Windows using Virtual PC**. That’s the best bet if you can invest the time needed to do the setup. Instructions are available [here](#). However, this will not be able to access your Windows file system. If that is important for you (like it is to me in most cases), please stick to Cygwin. On the other side, if you are testing code that will eventually run on a Linux machine, the VPC method is better.

## [Installing Knoppix within Windows](#)

By [Hardeep](#) • August 22, 2008 <http://blog.hardeep.name/computer/20080822/knoppix-in-windows/>

[Knoppix](#) is a [Linux](#) flavour that can run directly off a CD or DVD. All you have to do is boot off the CD and it starts running - it has all the basic applications like Internet Browser, Wordprocessing application etc. If you have FAT partitions, you will also be able to write to the disk.

However, I prefer booting into Windows and then running Knoppix as an application. (Although, I have a separate Debian installation as well for my main Linux needs.)

This can be done through Microsoft Virtual PC. I use the 2004 version downloadable [here](#). Also you need to have an ISO of the Knoppix CD. If you have a CD, you can make an ISO yourself, or download [here](#).

Start MS VPC 2004, create a virtual harddisk and load the ISO, then reboot the VPC machine. Now follow the instructions below:

1. Create a new VM with virtual HDD.
2. Attach the Knoppix ISO to the VM and reboot.
3. On the Linux prompt type: **linux install IGNORE\_CHECK=1 sudo knoppix\_installer**
4. On the menu that comes, select **3. Partition**
5. When it asks for it, select **Template 1**
6. Then back to main menu, select configure installation and accept all defaults
7. Then back to main menu, select start installation [now it will take time]
8. When the system boots, mouse won't be working so shut it down (but at least let it open the desktop first). Release the ISO and reboot VPC.
9. All file edits (required in steps below) will need to be done as root, by doing **su** first
10. Now, Grub will come up. On the first line, press **e**. Then on the next screen select the **kernel** line and press **e**. After this, add **i8042.noloop** at the end and press **Enter**. Press **Esc** and **Enter**.
11. Select the first option and boot.
12. After login, edit file `/boot/grub/menu.lst` and do the following for the first configuration only (one that says Default):
  1. Add **i8042.noloop** at the end of **kernel** line
  2. Remove the **savedefault** line since it causes problems later
13. Now edit file `/etc/sysconfig/desktop` and change **kdm** to **xdm**
14. Enter the command 'reboot' on the console
15. Let the system boot, login – mouse will work.
16. If the user config box comes up fine, else go to **Settings->Desktop setting wizard**. Now, select India in the first box and English-US in the second and press the first button.
17. Keep following the wizard. At the end select **launch KDE control center**.
18. Go to Regional and accessibility->Keyboard layout and bring up English layout
19. Done, take backup of the VHD file

Most of the steps should also apply if you want to install Knoppix on a normal HDD. Why we have to go this route - for example why mouse isn't working - beats me. Could be bugs - I have collected all this information from different places on the web, according to the problems I faced. If you know why we have to do this way, please place comments. Also, suggest improvements and let me know if this helped.

## MATHEMATICS

### The Spirograph

By [Hardeep](#) • November 17, 2008 <http://blog.hardeep.name/math/20081117/the-spirograph/>

I became a kid again and purchased a [Spirograph](#) set. The Spirograph is a mathematical toy, which you can use for drawing nice figures. In the simplest case it exists of a fixed circle, used as a template, and a smaller rolling circle with holes. The result of my experiment is at the end of this post.

Thereafter, I turned my attention to mathematical generation of the Spirograph figures, and to my surprise I was able to find a number of good resources on the net.

The parametric equations for a Spirograph are:

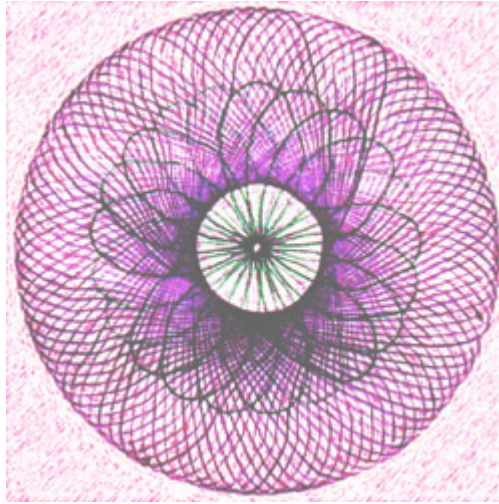
$$x(t) = (R+r)\cos(t) + p\cos((R+r)t/r)$$

$$y(t) = (R+r)\sin(t) + p\sin((R+r)t/r)$$

More explanation here:

<http://www.mathematische-basteleien.de/spirographs.htm>

An [applet](#), and [some data](#) to play with it can be [found here](#). However, what I found more useful and interesting is this page: <http://linuxgazette.net/133/luana.html>. It provides a Spirograph compiler (an awk script). This awk script takes a Spirograph specification and generates a gnuplot script to create the Spirograph. Go try it - its very interesting. If you are on windows, [read these](#) to help you get started on awk and gnuplot.



Spirograph

## [Value of Pi](#)

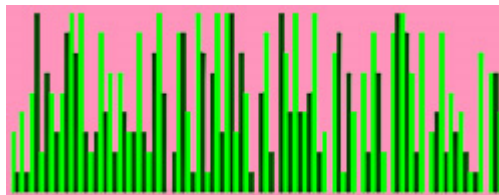
By [Hardeep](#) • July 25, 2008 <http://blog.hardeep.name/math/20080725/value-of-pi/>

$\pi$

(Pi - Source [en.wikipedia.org/wiki/File:Pi-symbol.svg](http://en.wikipedia.org/wiki/File:Pi-symbol.svg) )

Wikipedia defines Pi or p as “a mathematical constant which represents the ratio of any circle’s circumference to its diameter in Euclidean geometry”. As a kid I used to be interested in the calculation of Pi. The value of Pi, to 100 places of decimal is:

3.1415 9265 3589 7932 3846 2643 3832 7950 2884 1971 6939 9375 1058 2097 4944 5923 0781 6406 2862 0899 8628 0348 2534 2117 0664



Value to 100 places

I have calculated this using the Unix command [bc](#). The command for this is based on an identity (that I think is credited to Ramanujan) and is:



$$24*a(1/8)+8*a(1/57)+4*a(1/239)$$

where  $a$  stands for the arctangent function.

The formula is:

$$\pi = 24*\arctan(1/8)+8*\arctan(1/57)+4*\arctan(1/239)$$

First, load the bc language with associated library using “bc -l”. Then, set the scale to the number of digits using “scale=100”. Afterwards run the identity I gave above.

To experimentally calculate Pi experimentally, there are two ways: One using a random number generator and the other by physically measuring the circumference of a given circle.

Consider a square of length unity. Within that, a circle is drawn, having unity diameter. If a point is taken within the square at random, the probability of that point also lying within the circle is  $\pi*(1/2)*(1/2)$  which is  $\pi/4$ . Now, start taking points at random (x,y) and see if  $x*x+y*y \leq 1/4$  or not (if it is, then that means the point lies within the circle). Maintain the count of total number of points taken (t), and the number that fell within the circle(c). Now Pi can be calculated as:

$$\pi = 4*c/t$$

The other method is to take a circular bottle (measure the radius r) or tin and tie a thread around its circumference. Measure the circumference(c). Now Pi is  $c/2r$ .

The following two sentences contain the value of Pi: the number of letters in each word indicates the corresponding number in the value of Pi.

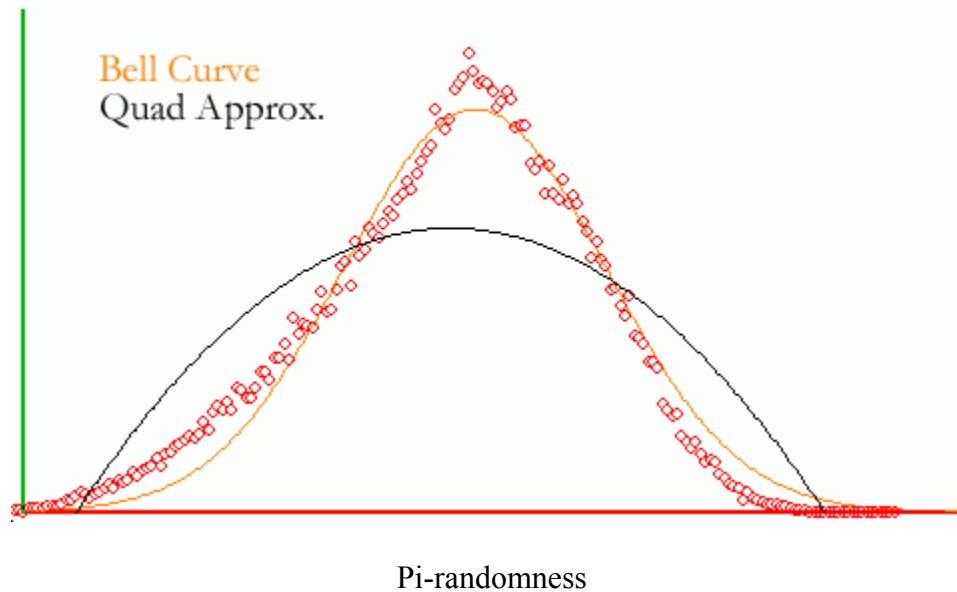
1. “May I have a large cup of coffee.”
2. “How I want a drink, alcoholic of course, after the heavy chapters involving quantum mechanics.”

This makes the value of Pi easy to remember.

Lastly, how useful are the digits of **Pi as a source of random numbers**? [Not bad](#), according to a study: “while sequences of digits from pi are indeed an acceptable source of randomness – often an important factor in data encryption and in solving certain physics problems – pi’s digit string does not always produce randomness as effectively as manufactured generators do”.

Always wanting to do my own thing, I downloaded the value of Pi to one million places from a website, split it into (x,y,z) coordinates, each having 5 digit precision. Here is the graph that got generated:





A bell curve, but could have been better - seems to mimic the results from the study.

## WORDPRESS BLOGGING

### [Adding a logo to the header](#)

By [Hardeep](#) • December 20, 2008 <http://blog.hardeep.name/computer/programming/20081220/logo-wordpress/>

Adding a logo to the header generated by Wordpress Default theme

When I started this blog, I did add categories '[Computers](#)' and '[Programming](#)'. However, **I made one decision**: I will not cover [HowTos](#) on Wordpress as part of this blog. The reason? There are at least two blogs for each Wordpress related topic on earth. Yes, I know about giving back to the community (Wordpress is free after all), but there are better ways in my opinion. The second reason was, have you ever noted someone **talking** on the Microphone **about** how to use **the Mike**? No, most people talk about other things, not about the Mike itself. Similarly, Wordpress is a medium. A lot of the Wordpress related blog posts are pilfered from elsewhere, even at times from Wordpress.org 😊

**Yet, this is my first (and hopefully the only) Wordpress [HowTo](#).** The reason is simple: I have come to realise that people come to blogs less to hear what is being said, and more learn how to speak their own thing. 😊 So, being human, and wanting to hike my readership I have joined the bandwagon. At least this isn't pilfered from anywhere.

Doing this is quite easy. Go to the folder `/wp-content/themes/default/images` under your Wordpress folder. In this folder, locate the file called `header-img.php`. You should see a line of code written like this:

```
//die;
header("Content-Type: image/jpeg");
```

Add two lines of code above this, making it look as below:

```
$hsim = imagecreatefrompng('logo.png');
imagecopy($im, $hsim, 70, 50, 0, 0, 100, 100);
```

```
//die;
header("Content-Type: image/jpeg");
imagejpeg($im, '', 92);
```

Here, logo.png is your logo image which is also lying the same folder. The numbers 70, 50 are the coordinates of upper left corner where you want your logo to appear - experiment a bit to find out what looks best for your logo. The numbers 100,100 are the width and height respectively of your logo image.

Have a look at how the result appears for me:



Forest fire  
November 14th, 2008

Logo Sample

Thats it! Easy - just try it and post any questions/queries/suggestions.

## [Hiding Wordpress categories](#)

By [Hardeep](#) • February 28, 2009 <http://blog.hardeep.name/computer/20090228/hiding-wordpress-categories/>



Photo by John Poyntz Tyler <http://www.flickr.com/photos/hname/3551117812/>

When I wrote [my first Wordpress related post](#), I admitted that I was only doing it to attract traffic and it would be my last post on the subject. However, I start again. This time around, however, I want to talk about something which isn't common knowledge and neither did I get any responses on the official Wordpress forum regarding this.

Suppose you do not want some of your posts to appear anywhere: not the homepage, not the RSS feeds, not the archives: nowhere. **However you DO want it to appear only when its linked to, as a single post on the page.** I regularly need to do this, because some part of the post is more like an 'addendum' or when including everything would make the post too long.

There is a standard solution available on the forums: creating a plugin and adding code to this effect:

```
function hs_cat_exclude($query)
{
```

```

if ($query->is_feed || $query->is_home || $query->is_archive ) {
$query->set('cat','-22');
}
return $query;
}

```

```
add_filter('pre_get_posts','hs_cat_exclude');
```

Here, 22 is the category number of the category I wanted to exclude.

This code works fine, but the moment you add `is_category` to the ‘if’ clause, it doesn’t work for the category page. This was perplexing to me, and I did not understand it. I spent a long time and then decided to dig deeper. I found out that the ‘wiring’ is faulty (this is what I believe). It can’t work like this for the category page. What is needed additionally is something like this:

```

function hs_cat_exclude_cat($where)
{
global $wp_query;
if ($wp_query->is_category) {
$where = $where . " AND NOT EXISTS(SELECT 1 FROM wp_term_relationships WHERE
wp_term_relationships.object_id=wp_posts.id AND wp_term_relationships.term_taxonomy_id='22')";
}
return $where;
}

```

```
add_filter('posts_where','hs_cat_exclude_cat');
```

So far so good. What I wanted over and above this though, is for the category to not even appear on the category widget. I tried to find a way to get this done through the plugin but it did not work. Ultimately I had to ‘hack’ one of the core files to achieve this. If anyone knows of a better way to accomplish this, please add a comment. The change is to `wp-includes/widgets.php`. Find the line of code that looks like:

```
$cat_args = array('orderby' => 'name', 'show_count' => $c, 'hierarchical' => $h);
```

added an ‘exclude’ clause like this:

```
$cat_args = array('orderby' => 'name', 'show_count' => $c, 'hierarchical' => $h, 'exclude' => 22);
```

Thats all there is to it.

# MANAGEMENT

## Top 10 considerations when preparing a software test plan

By [Hardeep Singh](http://blog.hardeep.name/computer/programming/20090515/test-planning-t10/) • May 15, 2009 <http://blog.hardeep.name/computer/programming/20090515/test-planning-t10/>



### -> Test the parts of the application that have changed since the last cycle / go live

This part of the test plan is very obvious: test the changes to the application. **Each change needs to be tested individually if possible, or as groups if the number of changes is large, and is known by the name regression testing.**

**For example**, if you added a new field called 'maximum pay by date' to the voucher batch interface, then you could test the interface for this - having both data with this date entered, and with this date set to blank.

There is nothing more to this one - its normally the facet of testing that **does receive the due focus** during testing.

### -> Test sampled parts of the application that have NOT changed

Now we come to something that **does NOT** receive the due focus. The parts of the application that remained unchanged. **No, you do not have to test ALL if it.** If you **can** test all of the application (especially with automated tools, as discussed below) - nothing like it. **However, at least test 10-15% of functionality that has not changed.**

For example - as discussed above - if you changed the voucher batch interface, then you can test the online voucher entry. Under the online voucher entry, test at least one scenario that has not changed.

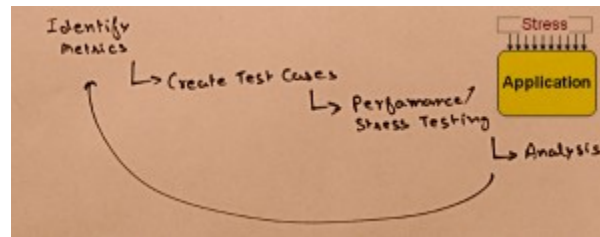
The rule of the thumb is that if in a module having 100 test cases, 40 have changed - then test those 40 that have changed, and test 6-10 of those 60 that have not changed.

### -> Look at it from the end users perspective: do one full cycle end to end

Next to include in the plan is something you can call integration testing: **if your application is about users entering vouchers and getting paid - perform this cycle as a user would do. Many times we IT folks test only our application - the one we are developing and forget the rest of the glue technology.** It falls into the category where we **want** to do it, yet are lazy at - so we find some short-cuts.

**Once I was asked** to carry out testing for a reconciliation report that had already been tested by the developers. I uploaded the same input twice, which ended up showing double on the final report. It turned out that the developer had missed this because he tested only on the basis of data that already existed in the system, and did not upload any new vouchers.

## -> Stress testing

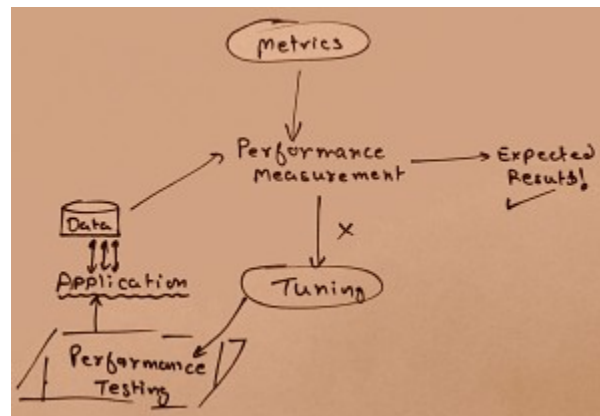


Stress testing should again be a very critical part of your test plan. **How many users are expected to use the application? during normal hours? during peak hours?** Plan for all such scenarios. Design the business process that would take place if the application does fail - the idea should be that the user's work doesn't get halted.

There are stress testing tools available both free and commercial that you can use to simulate users.

In one of my projects, a web application that was created for 800 users, failed under a load of 35. Increasing the number of processors, or the number of server boxes is not a guaranteed way of handling load on the application: the application has to be designed to support the load from the ground up, and tested suitably.

## -> Performance testing



How long does a file take to get processed? How long does the **user expect** it to take? How long it takes for the screen to open/save?

The user expectation part is sometimes ignored. Please go ask the users of your application now what their expectation is - or it might already be too late in terms of coding.

The developers might think if a process runs for one hour its good enough. However, the users might be needing to run it six times a day during the closing period. Hence one hour might not be fast enough. In such a scenario we had to run four parallel instances of a process to achieve the user specified timing.

## -> Concurrency testing

Can two different instances of the new process run together? The panel you just created: can it be used by two persons at the same time? Does it cause deadlocks at the database level if 100 instances of the process are run together?

Can two different versions of the application exist on the same machine?

These are the kind of questions that you ask yourself while working on the 'concurrency' aspect of test plan/execution.

A team of developers once needed to clone a process, and create slightly different functionality. However, it turned out that when both the processes were run together, 1 times in 10, one of the processes would fail. This was noted after go live 😊 Turned out the cause was incorrect use of the shared temporary tables by one of the processes.

[Read 'The what and the how of Deadlocks' also in this guide]

### **-> Unit test before Integration testing**

Our laziness at work again: we 'trust' our work and want to move directly to integration testing. Partially, the waterfall model of software development is also to blame here.

99% of the times, after the developer moves directly to integration testing - the very first test case for the application fails, and the developer comes back to the unit testing phase. 😊

Unit testing is a very critical part of your test plan - if you do it right, you will find hundreds of issues that will otherwise never get detected. Even not during integration testing.

Build 'driver modules' to iterate through all the 'ifs and whiles' that have been coded. Try out all avenues control can flow through.

### **-> Create test history**

Creation of a test history is as important as doing the testing. Being able to, at a later date, answer such questions as: 'what are cases we tested?', 'what are the problems we found?' etc is very helpful. Showing a clean slate (a 'pass' on all test cases) at the end of all our test iterations is not so helpful. In short, record the problems found, even though they may get corrected later on.

### **-> Automated testing**

Automated testing solutions can be a big help. It does not mean that all testing be delegated to the automated testing mechanism: but it can definitely be an add-on to your manual testing.

In changing the order entry functionality, use it to enter 1000 different orders. There are several solutions available (use google) that will record the user actions, and will repeat those actions later with different data.

At a very simple level, [AutoIt](#) is a great tool for automated data entry, and is free (GPL). Its very flexible and has a great library of functions built into its scripting language. I use it all the time, and not just for testing!

### **-> Code review**

While we focus on all these great ways of testing let us not forget our tried and tested workhorse: code review. Being humans, we are tempted to feel that by doing better testing (being easier to do) we can offset the need for a good code review, but there are hundreds of reasons to do code review.

There may be some program flows designed for rare situations which may never get tested. Code review in such a case will contribute ideas for such test cases. Documentation may not be in sync with the code, with the potential to make future changes difficult. There may be code improvements possible: for example, replacing an 'if condition' with a more specific check.

There are other things, depending on your scope you may also want to include them:

### **-> Knowledge transfer/competence testing**

### **-> Backup & recovery testing**

# Making scores comparable

By [Hardeep Singh](#) • May 27, 2009 <http://blog.hardeep.name/general/20090527/making-scores-comparable/>

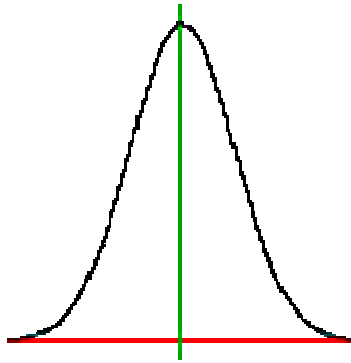
Assume that we have feedback on team members from two different project managers.

People->	U	V	W	X	Y	Z
Managers						
A	70	63	82	91	56	77
B	68	60	80	80	55	60

Can we say that W performs better in team A compared to team B? It looks like yes, he works better but analyse the scores a bit deeper. **Project manager B has rated all team members lower than manager A.** It may be that he is using tighter scoring.

In a different situation, it may be that two teams (of different people) have gone through two different tests, and we want to compare the people against others. Or, a university might want to compare people passed out in 2001 with those passed out in 2009.

**The question is, how do we compare people when the scores that we have do not use the same basis.**



Bell Curve (created using HSB Grapher)

The answer is: **normalization**. We fix the mean score, and the degree of deviation that we would like to see. For a 100 marks test, we may want 50 as the score and a 20% deviation. Now, we will compare this with the actual mean and the deviation of each of the sets, and modify the scores as needed. **Please refer to the [attached spreadsheet](#) which helps you do this.**

In the given example, A has a mean of 73, and a deviation of 13. B has a mean of 67 and a deviation of 11. **Let us bring both to a mean of 70 and a deviation of 15.**

People->	U	V	W	X	Y	Z
Managers						
A	66.3	58.1	80.4	91	49.9	74.5
B	71.2	60	87.9	87.9	53.1	60

So we note that A is indeed better in project A, but only slightly. Also from the initial figures we might have concluded that U is better in Project A, but actually the reverse is true.



Mathematically this process is called Normalization and is useful in fitting scores to a bell curve. Read more about it [here](#) if you are interested. However the spreadsheet attached is sufficient to get you started.

## **Benford's law - how to detect fraud through numbers**

By [Hardeep](#) • August 25, 2008 <http://blog.hardeep.name/math/20080825/benfords-law/>



Suppose you have some financial data - let us say all the vouchers paid by the company in a given month, and you want to run some tests to determine if there are any anomalies. For example, are the employees beating the approval process by entering say \$24.99 vouchers if the limit is \$25, or if any fraud is being committed. One way to do this is to use [Benford's law](#).

Benford's law states that in a given list of numbers generated naturally (for example stock prices or census figures), the probability of a number starting with 1 is 30.1%. The probability of a number starting with 2 is 17.6% and so on - it keeps decreasing as the numbers increase. The rationale behind it is explained as: it takes a 100% increase to take a number from 100 to 200. However, it takes only a 50% change to go from 200 to 300. 100% increase is more difficult to do (and thus has less probability of happening) than a 50% increase.

In this way, the probability of having a number starting with digit  $d$  is given by  $\log(1+1/d)$ , log to base 10. More information is available [here](#). Its usually extended to the first two digits for analysis in the real world.

[Download from here](#) a spreadsheet (called Numeric Truth) to carry out this analysis for you. All you have to do is to paste your data into the green cells. After that, on the first sheet it will show the results of first digit analysis, and on the second sheet, two digit analysis. Have a look at the graph, the variances for the individual digits, and the total variance. That should give you a starting point for your analysis/audit.

## **MISCELLANEOUS**

### **QR Codes**

By [Hardeep](#) • July 19, 2008 <http://blog.hardeep.name/computer/20080719/qr-codes/>

I came across QR codes recently. QR stands for Quick response. The coded message looks something like this:



<http://www.SeeingWithC.org>

In Japan, they are printed on business cards so that you can take a photograph of the code with your mobile phone, and use special software (also in the phone) to decode it. It makes your life simpler so that you don't have to feed the business card information into your phone contact book manually. However they can be used for almost anything - on ads to store contact information etc.

The QR code shown above **is my own web business card**. It has my name, the names of my websites and my contact email ID. You can go to <http://qrcode.kaywa.com/> to generate QR codes, and [download an application](#) to read them. Mobile phone applications are also available.

Why they attracted my attention was in connection to digital security: If you digitally sign a document using [GPG](#) for example, there is no way to reflect that on a printed version of the document. QR Codes present an easy way: include a QR code for a short summary of the document, digitally signed, into the document itself. To make it watertight, include a URL to the online version of the document, and a [hash](#) of the document. Makes sense?

The uses are endless: they can be added to ID cards, where on one side there is human-readable information, and on the other there is QR code, ready to be verified in case of suspected forgery, they can be added to marksheets digitally signed by the university - the list is endless. No softcopy of the marksheet needs to be provided is what makes this schema more interesting.

Here is a signed message from me:



Signed

No need to make them black and mundane - in fact you can superimpose your logo. The first reader to unravel the message in this QR Code will get a digitally signed certificate from me! My public key is [here](#). Post in comments.

## [Online judge](#)

By [Hardeep](#) • July 29, 2008 <http://blog.hardeep.name/computer/20080729/online-judge/>

Sometime back I came across [Sphere Online Judge](#) (SPOJ). A 'judge' is a mechanism through which you can verify your programming solutions.

The SPOJ contains a huge range of programming problems. You need to code a solution in one of the supported languages (from Unix shell scripting to C to Java to esoteric languages like B\*\*fcuk). When you submit the solution, the judge will execute your code with predetermined input and match against predetermined output. If all goes well, your solution will be accepted. Otherwise, it will say 'compilation error', or 'incorrect output' or 'time limit exceeded'. It will not tell you the input values or expected output, nor will it tell you the input for which your program is going wrong.

Its very interesting, the problems are very challenging and I gave some problems a try. At the end of it, you can see the report generated by the judge specific to you:

<http://www.spoj.pl/users/hardeeps/>

Also, you can see the submission history:

<http://www.spoj.pl/status/hardeeps/>

or, showcase a signed certificate that the judge provides:

<http://www.spoj.pl/status/hardeeps/signedlist/>

You can see that I have tried various languages, including Java, Bash (Unix shell) and Perl. I am yet to prove my programming skills in C. 😊

The result AC means that the solution is accepted. The certificate also shows that I was able to calculate PI correctly to 2500 places of decimal in less than 24 seconds - I used the identity described in [this blogpost](#).

I have tried to understand how to verify the digital signature on the certificate but have not been successful. The author informs me that a PHP function is being used to generate the certificate. It would be good if someone can get this working (please post in comments).

## Combating spam with GMAIL

By [Hardeep](#) • August 6, 2008 <http://blog.hardeep.name/computer/20080806/spam-gmail/>

I have been a fan of GMAIL ever since it was launched - the best feature is of course the quick search facility, and use it regularly. I learnt an interesting way to combat spam with GMAIL recently.

Suppose, the email ID you have is [abcd@gmail.com](#). You can add a '+' sign after abcd, and add whatever you want after that. For example, [abcd+friends@gmail.com](#) is also a valid email ID and the email sent to this address will reach you. When you are signing up with a service, let us say eBay, or a webforum and you fear the service provider will share/leak your ID, you can create a brand new email ID for them, for example [abcd+ebay@gmail.com](#). This way, you can setup special filters to automatically tag the emails, or if it starts getting too much spam you can setup filters to automatically trash all emails to this ID that do not come from ebay.com.

In fact, I later worked out an even better approach to combating spam with Gmail: The idea is to create two email IDs with Gmail (say abc@gmail.com and pqr@gmail.com) - one real and the other "hidden". The real one (abc) can be given to close friends who you trust not to spam. For all others do this: suppose you want to share your email ID with eBay.com and you dont trust eBay. Share this email ID with them: pqr+ebay@gmail.com. Setup a forwarder in 'pqr' account to forward all email sent to pqr+ebay@gmail.com to abc@gmail.com. All other email in the 'pqr' (for example sent to pqr@gmail.com) account can be set to be autodeleted. You get all your email in the abc@gmail.com account, without having to share it. More information on why this works is in [the comment](#).

**Why is this necessary?** Why cant we work the same way as I mentioned earlier? Here is why: Spammers are also smart. Suppose you create an account abc@gmail.com and share abc+ebay@gmail.com. Now, spammers can search there database searching for gmail addresses with a '+' in them and remove the '+' and the part following it. Now what happens? You start getting spam at your real email ID with no way to filter it. If we change our tactics slightly, with what I suggest above, in such a scenario spammers will send email to pqr@gmail.com, which isnt getting forwarded. Do try and let me know pitfalls. I know its a bit complex but should be worth it - have questions, post comments.

## Google releases Chrome

By [Hardeep](#) • September 5, 2008 <http://blog.hardeep.name/computer/20080905/google-chrome/>

Google [recently released](#) its new web-browser called [Google Chrome](#).

Microsoft had to undergo so many legal battles just because it released a web-browser for free. Today, there are so many free (and good) browsers available. Some credit should go to Microsoft for getting the ball rolling and providing

thought leadership to free (as in free of charge) software

I am yet to try this one out - waiting for the Linux version. Will post an update once I have tried it out. My current favourite is still [Firefox](#).

UPDATE June 08 2009: I tried Chrome - it has some nice features, but still cannot challenge Firefox. The main advantage is that it has an uncluttered interface, and leaves a lot of room for the 'content' - having minimal toolbars etc.

## Splitting a contact sheet

By [Hardeep](#) • September 24, 2008 <http://blog.hardeep.name/images/20080924/splitting-a-contact-sheet/>

When I received a CD containing contact sheets from my photographer, I was in a hurry to put the photos online. He had promised to send me another CD containing individual photos later, when the printing for the album was completed.

I tried to use normal software like Photoshop to do the trick, and looked for tools online. While there were a number of tools to create a contact sheet, there were none to take it apart. Strange.

After thinking, I was able to write a small Python program to do the trick. I am sharing it for the benefit of everyone, releasing it under [GNU GPL](#). [Download Python](#) and run it like: `python split.py Contact.jpg` (where `split.py` is a file containing the code below and `Contact.jpg` is the contact sheet).

Download [split.py](#) here.

Here, the contact sheet is 1000×800 pixels, and I want to split it into 5 parts horizontally and 4 parts vertically. These parameters need to be changed in the first four lines as per need.

I noticed that since the contact sheet was JPG, and the output was also JPG, it was causing degrading in quality. So, I changed the output format to BMP, and then reconverted back to JPG using other software.

## GIF vs JPEG

By [Hardeep](#) • January 10, 2009 <http://blog.hardeep.name/images/20090110/gif-vs-jpeg/>

There must have been times when you would have wondered what the difference is in pictures of type “GIF” (having file extension `.gif`) and those of type “JPEG” (having file extension `.jpg`). At other times you might have wondered which type to use for a picture you are uploading. Well, today I am going to solve the problem for you.

GIF stands for **Graphics Interchange Format**, and is more suited for computer generated graphics and images. This is because it can hold a maximum of 256 colors. If the picture you are saving as GIF has more colours, they will be “approximated”. Leaving aside this limitation, GIF is a wonderful format. You can mark some of the pixels as “**transparent**” which means that those pixels do not have a colour of their own, and will take up the background colour when the picture is displayed. This makes them merge easily with any background. In addition, you can create **animations** as well using GIF format.

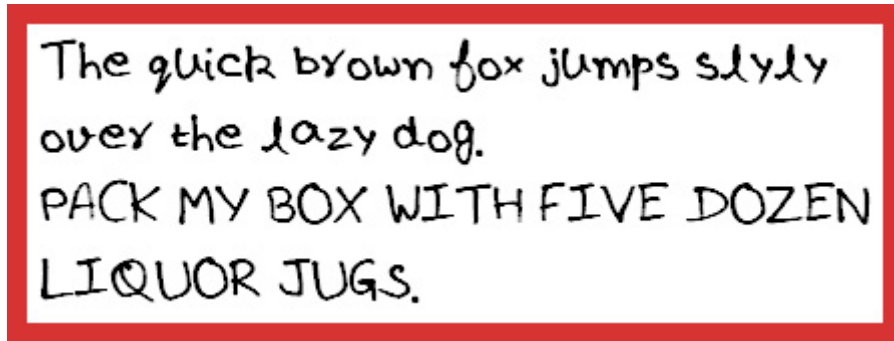
On the other hand, JPEG, standing for “**Joint Photographics Experts Group**” can store true 24 bit color, and is used more often for real life photographs. This is because real life photographs normally have more than 256 distinct colors. However, these support neither animations, nor transparent pixels.

The compression used by GIF is mostly lossless, while that used by JPEG is lossy - which means that a JPG image will lose some quality each time you save it, whereas a GIF will not.

## Handwriting font

By [Hardeep](#) • March 31, 2009 <http://blog.hardeep.name/computer/20090331/handwriting-font/>

Today I created a new font, from my own handwriting. Now if I need to write something - I can even type and print it in my own very handwriting!!! I think its cool for invitations etc.



This is what the font looks like

Jokes apart - you also can download it if you need it for any reason: [here](#).

Please let me know how you used it.

## The world of computers: vision 2020

By [Hardeep Singh](#) • May 16, 2009 <http://blog.hardeep.name/computer/20090516/vision2020/>



What would the shape of computing be 12-15 years from now? Here is where I think we will be:

**My wrist watch will have my computer.** When I reach office, I will place myself in front of a 'dumb' terminal - a monitor, a keyboard and a mouse. Embedded into the keyboard will be a smart card which will talk to my wrist watch (without cables). **I will use a remote log-on software to connect to the computer inside the wrist watch - all applications will already be installed on the wrist watch and I will use them.** It will also be possible to use the wrist watch as a pen drive of today. So all the data on the hard disk of this computer will be available in two ways: the remote log-on (which will also enable the use of installed applications), and USB (that is, minus the capability to use apps).

At home (and everywhere else), I will have a similar dumb terminal.

**Microsoft** will be dead - opensource (and [portable](#)) software like OpenOffice, and AbiWord will have caught up in terms of functionality. For profit firms of 2020 will provide support (and contribute to the enhancement to) GPL software.

**Google** will be going, but its offering of (office and other) applications as an online subscription (which will have become paid by then) will not be doing very well. People want to collaborate, but not at the expense of being tied down.

**Electronics commerce** will still have identity fraud 😊 Sorry guys. However, the total volume digitally traded will be rising steadily.

**Digital signatures** would be much more easier to use, and transparent to the uninitiated user. However, it will not be free from its own share of frauds.

**Operating systems** will be very different from today: there will be no device drivers. Every device will be plug & play, and will use universal drivers. Linux will be the defacto standard.

You have some more ideas? Please feel free to share.





# About the author...

The author **Hardeep Singh** is a software project manager by profession, MBA (Finance) by education and a photographer by hobby. He holds the ISO 9000 auditor and ITIL (Foundation) certifications, and is a Linux enthusiast.

He blogs at **Digital Wealth** <http://blog.hardeep.name> on a range of topics from Computers, to Finance, to Photography and Religion. The name Digital Wealth represents knowledge in a digital form.

This magazine is licensed under Creative Commons Attribution - Non Commercial - No Derivative Works 2.5 India License. For the purpose of this license, this entire magazine is a single entity and cannot be taken apart. Publishing it verbatim either printed or in PDF format is sufficient attribution and no other action is needed towards this.

The terms of this license are available online here:

<http://creativecommons.org/licenses/by-nc-nd/2.5/in/>