

*Workshop on  
Digital Libraries: Theory and Practice  
March, 2003  
DRTC, Bangalore*

**Paper: T**

## **Installing and Running Harvest on Linux Platform**

**A.R.D. Prasad**

Documentation Research and Training Centre

Indian Statistical Institute

Bangalore – 560059

email: [ardprasad@hotmail.com](mailto:ardprasad@hotmail.com)

### **Abstract**

*This paper presents a very brief introduction to installing and running Harvest system. The various steps involved in installing and running are given with simple explanation and it is no substitution to the Harvest Manual.*

## 1. STEPS IN CONFIGURING HARVEST

This tutorial explains a basic approach to install Harvest on Linux platform and to be more specific RedHat Linux 7.3. However the same should work on other Linux variants like Debian, Mandrake, etc. with little or no modifications. The directories used in the tutorial are only suggestions, if you wish to load Harvest on to different directories you are free to do so. For system requirement please refer the Harvest manual

*Step 1:* Create a directory with name `/home/harvest` under your home directory. Change your working directory to that directory.

*Step 2:* Download Harvest file `harvest-1.9.4.tar.gz` from the following URL. You can also download the documentation file `harvest-1.7.15-docs.tar.gz`

**<http://prdownloads.sourceforge.net/harvest/>**

*Step 3:* Once you are `harvest` directory and downloaded the files mentioned in step 2, use the following command to untar and unzip the harvest file, using the following command

**`tar zxvf harvest-1.9.4.tar.gz`**

*Step 4:* Change the working directory to `harvest-1.9.4`

*Step 5:* Read the INSTALL file, for detailed installation instructions. As Harvest is not distributed as binary file, we have configure and compile the Harvest System. Issue the following command to configure Harvest. (You can enter `./configure --help` for complete list of arguments for configure command).

**`./configure --prefix=/home/harvest`**

The above command assumes that you wish to install Harvest in `'/home/harvest'` directory. It takes a couple of minutes to run the configure command. The system outputs lots of messages about the availability of various components required to compile Harvest and also creates the necessary make files.

*Step 6:* In order to compile Harvest, issue the following command

**`make`**

It takes a longer time to compile. You will see lots of compiler messages and if everything is compiled, you should not get any error messages.

*Step 7:* Once the Harvest sources programs are compiled, the next step is to install Harvest with the following command

**`make install`**

The above command installs harvest in `'/home/harvest'`

*Step 8:* Go to `'/home/harvest'` with the following command

**`cd /home/harvest`**

You will find the following files and directories

- Bin
- Brokers
- Cgi-bin
- Gatherers
- Lib
- Man
- RunHarvest

Share

Except the 'RunHarvest' (command) others are directories.

## 2. INTERACTIVE SESSION OF RUNNING HARVEST

*Step 8:* Until now, we have only configured, compiled and installed Harvest. In this step, we run Harvest using the following command

***./RunHarvest***

*Step 9:* The following questions are asked when you run the interactive session of setting up Harvest.

**Q:** Do you want to continue? [Yes]

**Ans:** just press the enter key (The default answer will be shown in the square brackets. When you press the enter key, the default answer will be chosen).

**Q:** Do you want to continue? [Yes]

**Ans:** just press the enter key. Both the above questions are presented by giving an information about Harvest to have better understanding of the system.

**Q:** On which host does your WWW server run ? [drtc.isibang.ac.in]:

**Ans:** Harvest normally picks up your server name. You can simply press the return key.

**Q:** On which port does your WWW server run? [80]

**Ans:** Mostly, web servers are run on port 80, you can simply press the enter key. However, some times web servers are configured on different ports, to check and make sure check the /etc/services file and look for the line beginning with http. Normally, it will look like the following line

http 80/udp

http 90/tcp

**Q:** Select a standard configuration

We offer 3 standard configurations

- 1) Index your entire WWW site
- 2) Index an entire WWW site (or sites)
- 3) Index selected parts of WWW, FTP or Gopher sites

Please select a configuration: [1]

**A:** The options indicate the choices available. For our preliminary test, press enter key i.e. to index your own web site. Indexing web sites located on other machines on Internet much longer time and heavily depends on the network speed. However the second option is very useful in developing subject/information gateways.

**Q:** Enter a short description of this Harvest server:

**A:** Whatever is entered will be used as the heading of your Harvest server. Or you can simply press the return key.

**Q:** Enter a one word description of the Harvest server:

**A:** the default is nothing but the domain name of your web server. However, you can enter some mnemonic name (one word) of your Harvest. You can create any number of Harvest servers on a single machine. You can choose a right word which represents the contents of your server.

Normally, a directory will be created under which all the index files are placed.

**Q:** Where do you want to install the Gatherer?

**A:** Choose the default option by pressing the enter key.

**Q:** On which port should the gatherer run? [8500]

**A:** Make sure from the '/etc/services' file that port 8500 is not used by any other services, before you press the enter key.

**Q:** Where do you want to install the Broker?

**A:** Choose the default option by pressing the enter key.

**Q:** On which port should the broker run? [8501]

Q: Enter a password for Broker administrative command

A: You know the answer

Q: Would you like to edit the Gthers' workload specification? [no]

A: You can choose the default answer i.e. [no]. This takes lot of time and even more when you try to gather information from other sites on the Internet.

**Final Message:**

You will get the following messages.

Your Harvest servers are now running. To access them refer to

**<http://drtc.isibang.ac.in:80/Harvest/brokers/drtc/summary.html>**

### **3. APACHE WEB SERVER CONFIGURATION**

The last step is to configure the Apache Web server. You require to have system administration password (root password). Or you can ask the system administration to make the necessary changes to http.conf file.

```
ScriptAlias /Harvest/cgi-bin/ /home/harvest/cgi-bin/
```

```
Alias /Harvest /home/harvest
```

Once the above two lines are added in http.conf file. You have to restart the web server by enter the following command

```
/etc/rc.d/init.d/httpd restart
```

Lastly open a web browser enter either of the following URLs

**<http://drtc.isibang.ac.in/Harvest/brokers/drtc/summary.html>**

OR **<http://drtc.isibang.ac.in/Harvest/brokers/drtc>**

### **4. TROUBLE SHOOTING**

For some strange reason, sometimes, you may not see any results. In such a case, enter the following URL

**<http://drtc.isibang.ac.in/Harvest/brokers/faculty/admin/admin.html>**

Once the relevant screen is displayed, enter the password and click the button "issue command". It attempts to rebuild the collection, so that the search results are displayed next time, when you attempt to search.

### **5. CONCLUSION**

This article presents only the basic approach. However, to utilize the full power of Harvest, you are strongly advised to refer the documentation of Harvest.