calamus, Fotolia

Breaking through the backup barrier

SMOOTH OPERATOR

BackupPC handles backups over the network for a ranges of platforms. Find out more about this user-friendly, configurable, high-performance open source backup system. **BY DAVID NALLEY**

Network backup platforms are often unwieldy, partly because of the complexities of scheduling logic and media management. User friendliness can be hard to find in an enterprise-ready backup system. The BackupPC project [1] fills the backup niche elegantly, handling backups over the network for a plethora of platforms and transports.

BackupPC follows the Unix tradition of small programs that perform a single task very well. Like other classic Unix utilities, BackupPC leverages the power of other applications instead of trying to reinvent the wheel. BackupPC supports several protocols for both Windows and Unix-like clients – from rsync and SMB/CIFS, to tar and rsync tunneling over SSH. The focus is on efficient scheduling and a user-friendly restore process.

BackupPC has an active user community with mailing lists and a user-generated wiki, and the project is still led by the original primary author, Craig Barratt. Although the tool has been around since 2001 and is relatively mature, the latest version – BackupPC 3.1.0 – seems to be reaching new users.

Benefits

One of the defining features of BackupPC is data de-duplication. In a traditional backup system, having multiple backups of files that haven't changed in more than one full backup interval requires storing the same information more than once. The problem is only compounded when you back up multiple computers – particularly if they are end-user machines that might be on the same circulation list for memos, spreadsheet, and other common documents. BackupPC addresses this problem with a two-tier check. The first check locates files with the same names and hashes the files to see whether they are identical. If it determines that the files are the same, it moves a single copy of the file to the "pool" and creates hard links to each instance of the file in the backup set. The results are surprising: In the first test I ran on eight machines (performing uncompressed backups and retention of two full backups and six incrementals), my total data store was ~1TB, but BackupPC's data de-duplication brought the actual size on disk to ~675GB.

BackupPC also offers several nice scheduling features, such as the ability to prioritize backups. By default, BackupPC wakes hourly and identifies any computer that hasn't completed a backup within the specified interval. Also, it checks to see which machines are on the network, and after combining these two lists, BackupPC prioritizes the list of available hosts on the basis of time since the last backup. Other factors can also influence this priority list. For instance, a machine that is on the network 24 hours a day is generally preempted by a machine with a more sporadic network presence record.

My favorite feature of BackupPC is that end users can initiate and perform their own restores without the interaction of the backup operator or system administrator. If you have been involved in backups on any scale, you know that handling restores of a lost or mangled file is time consuming. If the user needs to find a specific version of the file, the restore process can grow into a multihour effort. BackupPC offers a friendly web interface that provides a directory and file tree for each backup. Users can select a single file or multiple files in the tree, and BackupPC will restore these files without the need for a system administrator. BackupPC even checks for whether the user has the necessary access permissions to view the file before beginning the restore.

Users also have some control over when to start a backup (full or incremental) or whether to remove their machines from the backup list for a number of hours.

Installation

Installation of BackupPC is relatively painless because it's included in most mainstream distribution package repositories. However, sometimes it doesn't include the latest available code or has some special installation requirements, so I'll cover installation from source. If you used your distribution's package for BackupPC, skip ahead to the Configuration section.

Before working on the installation, you must consider disk space and how it is set up. Because BackupPC handles de-duplication by creating hard links from the file location in the directory structure of the backups to the pool where duplicated files are actually stored, the backup store must be on a single filesystem. This doesn't mean that you can't use LVM, software RAID, or hardware RAID to combine multiple disks, but you can only use a single filesystem to hold the store.

BackupPC tests to ensure that it can create these hard links at each startup. You'll need to know the mount point of this filesystem during installation.

The next two steps are really one-liners in the console and consist of creating a user for BackupPC to run as and installing the software prerequisites. Of course, this assumes that you have httpd already installed and configured for your server:

```
# adduser backuppc
# yum install ⮐
perl-Compress-Zlib ⮐
perl-Archive-Zip ⮐
perl-XML-RSS ⮐
perl-File-RsyncP
```

After the prerequisites are out of the way, you can grab the source [1] and uncompress it:

```
$ tar -zxvf ⮐
BackupPC-3.1.0.tar.gz
$ cd BackupPC-3.1.0
$ su -c "perl ./configure.pl"
```

This launches the installer, which performs the basic configuration and installation of BackupPC. The default answers are fine, with a few exceptions. The data directory should be the mount point of the filesystem for the backup pool (e.g., */data/BackupPC* or a subdirectory therein). Also, you might need to enter the correct path for the CGI bin directory (e.g., */var/www/cgi-bin/*).

So that BackupPC will start automatically, you must add init scripts to your system. In the *init.d* subdirectory, you will find init scripts for a variety of distributions.

Copy this to */etc/init.d*, set it to start on boot, and then start the daemon:

```
$ su -c ⮐
"cp linux-backuppc ⮐
/etc/init.d/backuppc"
$ su -c "chkconfig ⮐
--add backuppc"
$ su -c "chkconfig ⮐
--level 345 backuppc on"
$ su -c "chkconfig ⮐
--list backuppc"
$ su -c "service ⮐
backuppc start"
```

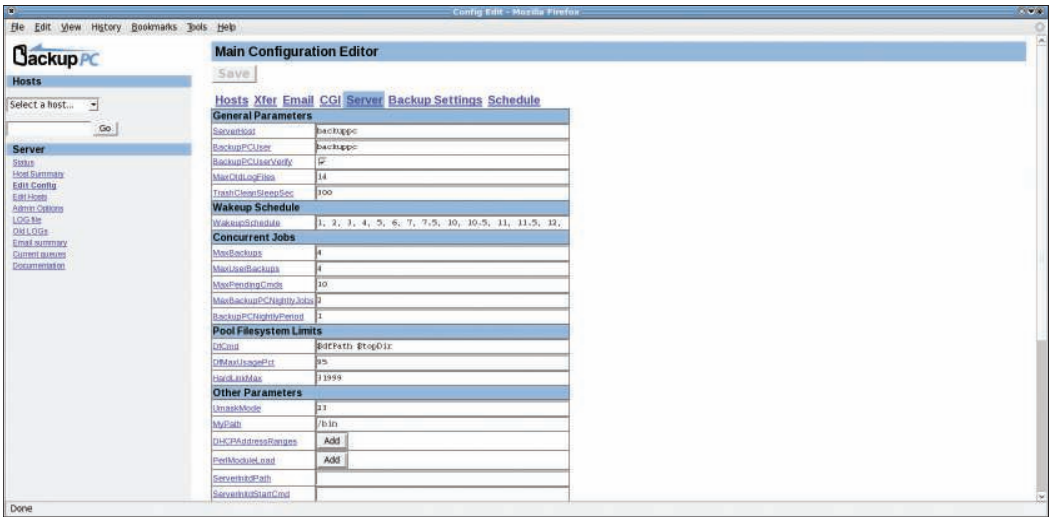
Configuration

Although the installation process handles the basic configuration elements, other options are available via the web interface or command line.

BackupPC configuration is contained in two files under */etc/BackupPC*: *hosts* details the identity of the hosts to be backed up, and *config.pl* controls the server configuration.

The hosts file lists the hostnames to be backed up and the authorized users for that machine:

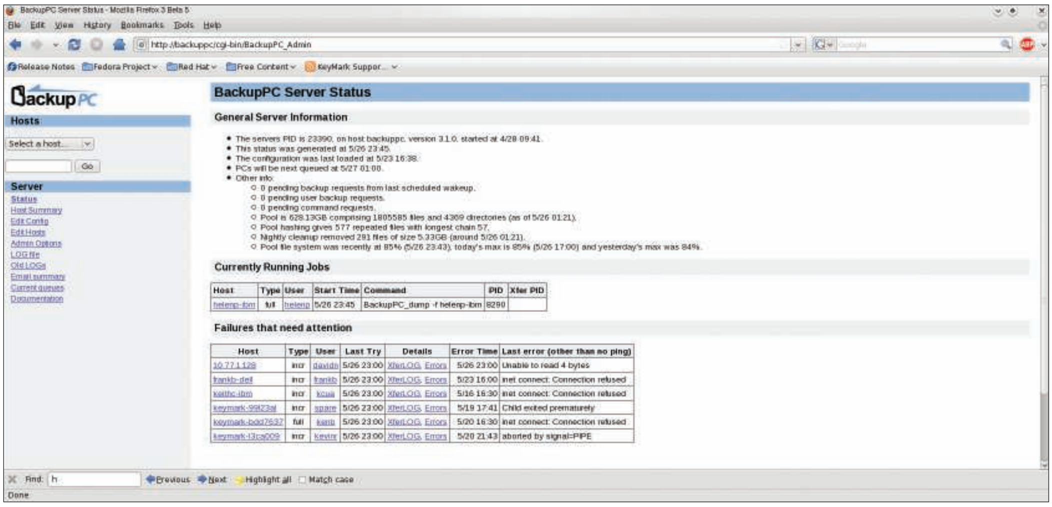
```
host   dhcp  user  moreUsers ⮐
# <--- do not edit this line
nalleyt61  0     david  ⮐
# <--- example static ⮐
IP host entry
host2    1    bill    jeff,fred  ⮐
# <--- example DHCP host entry
```



Main Configuration Editor

General Parameters	
ServerHost	backuppc
BackupPCUser	backuppc
BackupPCUserVerify	☑
MaxOldLogFiles	14
TrashCleanSleepSec	300
Wakeup Schedule	
WakeupSchedule	1, 2, 3, 4, 5, 6, 7, 7.5, 10, 10.5, 11, 11.5, 12,
Concurrent Jobs	
MaxBackups	4
MaxUserBackups	4
MaxPendingCmds	10
MaxBackupPCNightlyJobs	2
BackupPCNightlyPeriod	1
Pool Filesystem Limits	
DfCmd	$dfPath $topDir
DfMaxUsagePct	95
HardLinkMax	31999
Other Parameters	
UmaskMode	23
MyPath	/bin
DHCPAddressRanges	Add
PerlModuleLoad	Add
ServerInitdPath	
ServerInitdStartCmd	

Figure 1: In the Configuration Editor, you can adjust the server configuration settings.



BackupPC Server Status - Mozilla Firefox 3 Beta 5

http://backuppc/cgi-bin/BackupPC_Admin

BackupPC Server Status

General Server Information

- The servers PID is 23390, on host backuppc, version 3.1.0, started at 4/28 09:41.
- This status was generated at 5/26 23:45.
- The configuration was last loaded at 5/23 16:38.
- PCs will be next queued at 5/27 01:00.
- Other info:
  - 0 pending backup requests from last scheduled wakeup.
  - 0 pending user backup requests.
  - 0 pending command requests.
  - Pool is 628.13GB comprising 1805585 files and 4369 directories (as of 5/26 01:21).
  - Pool hashing gives 577 repeated files with longest chain 57.
  - Nightly cleanup removed 281 files of size 5.33GB (around 5/26 01:21).
  - Pool file system was recently at 85% (5/26 23:43), today's max is 85% (5/26 17:00) and yesterday's max was 84%.

Currently Running Jobs

Host	Type	User	Start Time	Command	PID	Xfer PID
helenp-ibm	full	helenp	5/26 23:45	BackupPC_dump -f helenp-ibm	8290	

Failures that need attention

Host	Type	User	Last Try	Details	Error Time	Last error (other than no ping)
10.77.1.128	incr	davidn	5/26 23:00	XferLOG, Errors	5/26 23:00	Unable to read 4 bytes
frankb-dell	incr	frankb	5/26 23:00	XferLOG, Errors	5/23 16:00	inet connect: Connection refused
keithc-ibm	incr	kcua	5/26 23:00	XferLOG, Errors	5/16 16:30	inet connect: Connection refused
keymark-99f23al	incr	spare	5/26 23:00	XferLOG, Errors	5/19 17:41	Child exited prematurely
keymark-bdd7637	full	kerib	5/26 23:00	XferLOG, Errors	5/20 16:30	inet connect: Connection refused
keymark-l3ca009	incr	kevinr	5/26 23:00	XferLOG, Errors	5/20 21:43	aborted by signal=PIPE

Find: h

Done

Figure 2: Check the status of currently running backups and failures.

When I cover authenticating to the web interface, I'll explain authorized users more, but the vital points are the hostname and the DHCP setting. If your machine gets its address via DHCP, you still want to use *0* for the DHCP setting, which tells BackupPC to use DNS to find the host. Setting this value to *1* tells BackupPC only to use nmblookup to query for the host address via NetBIOS.

The default *config.pl* is configured to wake up every hour to look for hosts to back up, do a full backup approximately every 7 days, and do an incremental backup every day (Figure 1).

You can adjust these – and other – settings. The manual and the config file provide detail about each option, but I will only look at the minimum options that must be configured to start backups on either Windows or Linux. Also, it's important to remember that you can make modifications on a per-machine basis, too.

One thing to set up is the admin user and how backups will be transferred in your environment (see Listing 1). However, I don't advocate the use of root as the backup user; instead, I suggest that you use a low-privileged account and set up sudo so that rsync is accessible. As the backuppc user, you'll also need to log in to the client machine via SSH so that it becomes a known host.

With the use of visudo, set up the following line in */etc/sudoers* on the client machine,

```
backuppc ALL=NOPASSWD: /usr/bin/rsync
```

then modify the command arguments so that it uses sudo to call rsync:

```
$Conf{RsyncClientCmd} = '$sshPath -l backup $host nice -n 19 sudo /path/to/rsyncSend $argList+';
```

Although you shouldn't limit yourself to just these configuration options, setting these items at a minimum will take care of backing up Windows machines or Linux machines with Samba shares exposed.

Although you can configure a number of other things, such as file/directory exclusions and compression levels, the last required item is configuring the web interface. The installation automatically installed the web interface, but you need to set up authentication for it, and you need a way to authenticate the users in the hosts file and the admin users. Because you are using Apache to provide authentication, you have a variety of ways to authenticate. For instance, you could use LDAP, Active Directory, basic authentication, or anything else Apache supports.

Listing 1: config.pl Options

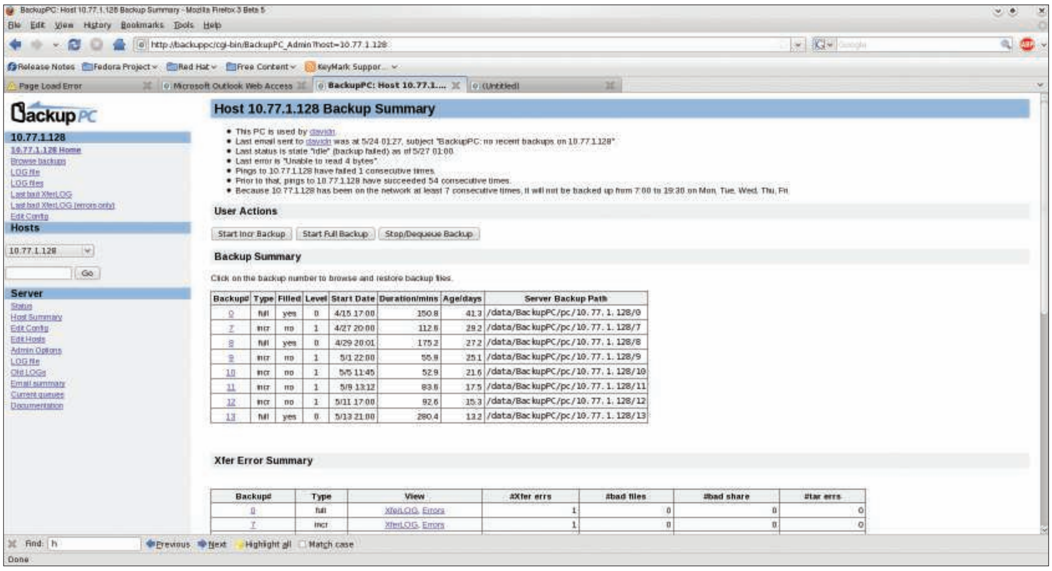
```
$Conf{CgiAdminUsers}  = 'ke4qqq,bill';  #sets the admin users to bill and ke4qqq
$Conf{XferMethod} = 'smb';                  #sets the default transfer method to smb (other options:
   rsync, rsyncd, or tar).
$Conf{SmbShareName} = 'C$';               #the default smb share name, in this case the admin share for the
   C drive
$Conf{SmbShareUserName} = 'david';    #the username for the smb share
$Conf{SmbSharePasswd} = 'password';   #the password for the smb share

#If you are backing up mainly Linux machines, the following settings are also appropriate.

$Conf{XferMethod} = 'rsync'                   #sets the default transfer method to rsync
$Conf{RsyncShareName} = '/';          #specifies the directory you want to backup.
$Conf{RsyncClientRestoreCmd} = '$sshPath -q -x -l root $host $rsyncPath $argList+'; # tells BackupPC to
   use ssh and
# then rsync. You'll need to set up
# ssh keys for the backuppc user
# and then copy them over.
```

Although Barratt's documentation delves into setting up LDAP authentication, among others, I'll focus on basic digest authentication, which requires you to add the section shown in Listing 2 to *httpd.conf*. Then, you'll want to run:

```
htpasswd -c ➘
/etc/httpd/conf/ ➘
passwd ke4qqq
htpasswd ➘
/etc/httpd/conf/ ➘
passwd bill
```

Note that the *-c* switch is only used when initially creating the password file; omit it for each subsequent user. This will prompt you for ke4qqq's password in the file.

After reloading httpd and starting BackupPC, you should be able to launch a browser and point it to *http://backuppcserver/cgi-bin/BackupPC_Admin/ authenticate* as a user you created and gain access to the web interface. If you aren't an admin user, you'll only have access to machines on which you are listed as the user in the host file.



Host 10.77.1.128 Backup Summary

- This PC is used by davidn.
- Last email sent to davidn was at 5/24 01:27, subject "BackupPC: no recent backups on 10.77.1.128"
- Last status is state "idle" (backup failed) as of 5/27 01:00.
- Last error is "Unable to read 4 bytes".
- Pings to 10.77.1.128 have failed 1 consecutive times.
- Prior to that, pings to 10.77.1.128 have succeeded 54 consecutive times.
- Because 10.77.1.128 has been on the network at least 7 consecutive times, it will not be backed up from 7:00 to 19:30 on Mon, Tue, Wed, Thu, Fri.

User Actions

Start Incr Backup | Start Full Backup | Stop/Dequeue Backup

Backup Summary

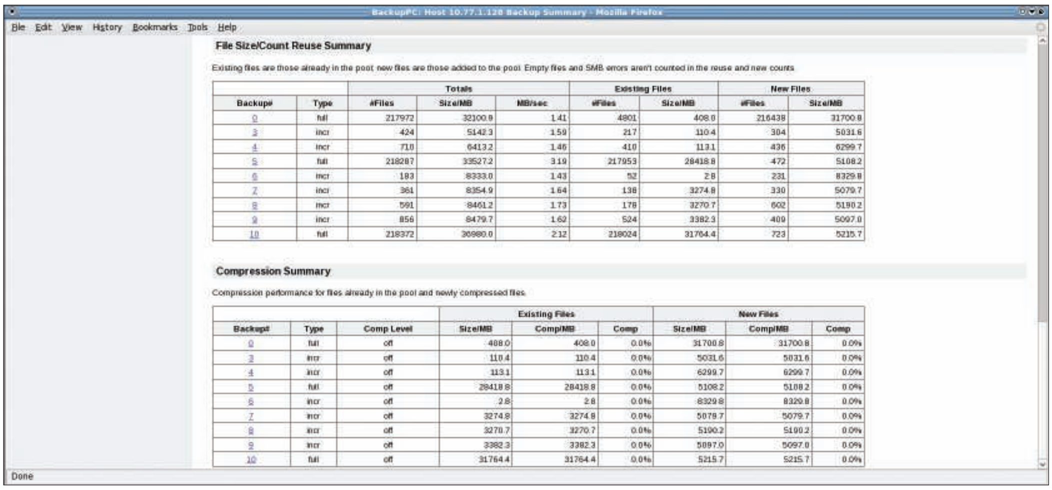
Click on the backup number to browse and restore backup files.

Backup#	Type	Filled	Level	Start Date	Duration/mins	Age/days	Server Backup Path
0	full	yes	0	4/15 17:00	150.8	41.3	/data/BackupPC/pc/10.77.1.128/0
7	incr	no	1	4/27 20:00	112.6	29.2	/data/BackupPC/pc/10.77.1.128/7
8	full	yes	0	4/29 20:01	175.2	27.2	/data/BackupPC/pc/10.77.1.128/8
9	incr	no	1	5/1 22:00	55.8	25.1	/data/BackupPC/pc/10.77.1.128/9
10	incr	no	1	5/5 11:45	52.9	21.6	/data/BackupPC/pc/10.77.1.128/10
11	incr	no	1	5/9 13:12	83.6	17.5	/data/BackupPC/pc/10.77.1.128/11
12	incr	no	1	5/11 17:00	92.6	15.3	/data/BackupPC/pc/10.77.1.128/12
13	full	yes	0	5/13 21:00	280.4	13.2	/data/BackupPC/pc/10.77.1.128/13

Xfer Error Summary

Backup#	Type	View	#Xfer errs	#bad files	#bad share	#tar errs
0	full	XferLOG, Errors	1	0	0	0
7	incr	XferLOG, Errors	1	0	0	0

Figure 3: The status page lets you start, stop, and dequeue backups ...


BackupPC: Host 10.77.1.128 Backup Summary - Mozilla Firefox
File Edit View History Bookmarks Tools Help
Done

File Size/Count Reuse Summary

Existing files are those already in the pool; new files are those added to the pool. Empty files and SMB errors aren't counted in the reuse and new counts.

		Totals			Existing Files		New Files	
Backup#	Type	#Files	Size/MB	MB/sec	#Files	Size/MB	#Files	Size/MB
0	full	217972	32100.8	1.41	4801	408.0	216438	31700.8
3	incr	424	5142.3	1.59	217	110.4	304	5031.6
4	incr	710	6413.2	1.46	410	113.1	436	6299.7
5	full	218287	33527.2	3.19	217953	28418.8	472	5108.2
6	incr	183	8333.0	1.43	52	2.8	231	8329.8
7	incr	361	8354.9	1.64	138	3274.8	330	5079.7
8	incr	591	8461.2	1.73	178	3270.7	602	5190.2
9	incr	856	8479.7	1.62	524	3382.3	409	5097.0
10	full	218372	36980.0	2.12	218024	31764.4	723	5215.7

Compression Summary

Compression performance for files already in the pool and newly compressed files.

			Existing Files			New Files		
Backup#	Type	Comp Level	Size/MB	Comp/MB	Comp	Size/MB	Comp/MB	Comp
0	full	off	408.0	408.0	0.0%	31700.8	31700.8	0.0%
3	incr	off	110.4	110.4	0.0%	5031.6	5031.6	0.0%
4	incr	off	113.1	113.1	0.0%	6299.7	6299.7	0.0%
5	full	off	28418.8	28418.8	0.0%	5108.2	5108.2	0.0%
6	incr	off	2.8	2.8	0.0%	8329.8	8329.8	0.0%
7	incr	off	3274.8	3274.8	0.0%	5079.7	5079.7	0.0%
8	incr	off	3270.7	3270.7	0.0%	5190.2	5190.2	0.0%
9	incr	off	3382.3	3382.3	0.0%	5097.0	5097.0	0.0%
10	full	off	31764.4	31764.4	0.0%	5215.7	5215.7	0.0%

Figure 4: ... and you can check statistics about each one.

Web Interface

After the server is set up, you'll want to wander around the user interface a bit. On most installations, the URL should be: *http://yourbackuppcserver/cgi-bin/BackupPC_Admin*. This will bring you to the Server Status page (Figure 2). If you are one of the administrative users, you will see all of the currently running backups and any failures that need attention. Most of these failures will be caused by users stopping the backup by shutting off their machines. Also, you can see statistics about the server.

One of the nice things about the interface is that virtually all of the references to backups or hosts are clickable and will carry you either to the status page or the browse page for the backup. You'll also see a link to the documentation, which is served up locally.

Listing 2: Modifying httpd.conf

```
<Location /cgi-bin/BackupPC_Admin>    # <--- change path as needed
       AuthType Basic
       AuthName "BackupPC login"
       AuthUserFile /etc/http/conf/passwd    # <--- change path as needed
       require valid-user
</Location>
```

On the left-hand side, you will notice all of the navigation links for the user interface. The drop-down menu at the top presents a list of hosts, and the search box below lets you search for hosts by name. After you select one of the hosts, it will permit you to go to its status page, which shows you a list of all of the backups that have completed along with statistics about each one, such as size and when the machine was backed up (Figures 3 and 4). On the status page for each machine, you can also start and stop backups and dequeue a machine for a given number of hours. The end user assigned to the machine in the hosts file is also able to start, stop, and dequeue backups.

Clicking on the list of backups presented on each host's page will take you to the restore interface. On the Backup Browse page, you'll find a navigation tree along the left side for the filesystem and a list of files to the right (Figure 5). Within this interface, you can click on a single file and download the file directly in your browser, or you can select multiple files by marking the checkbox and then clicking *Download selected files*.

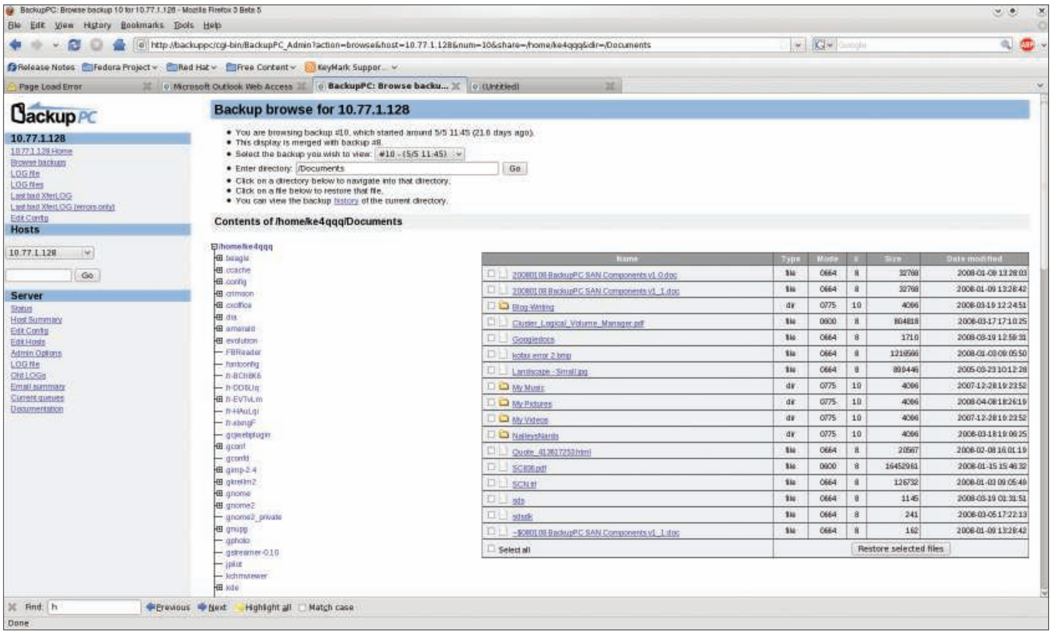
On the next page, you can specify the files to be compressed or whether you want BackupPC to restore the files directly to your machine via the same transfer method that is used to back the files up. Note that the transfer method will need write access to the filesystem (Figure 6).

One other page frequently accessed by administrators is the Host Summary page (Figure 7), which provides a color-coded table of all of the hosts configured for the machine. In addition to the colors, which indicate current status, you can see the size of the last full and incremental backups, the speed of the backup, time since last backup, and last full backup. This gives you a quick bird's eye view of all the machines being backed up and quickly allows you to locate any trends or persistent problems.

BackupPC provides a comprehensive backup solution that is also user friendly. If you are looking to deploy a backup solution – particularly for end-user machines – you should certainly consider BackupPC. ■

INFO

[1] BackupPC: *http://backuppc.sourceforge.net/*



http://backuppc/cgi-bin/BackupPC_Admin?action=browse&host=10.77.1.128&num=10&share=/home/ke4qqq&dir=/Documents

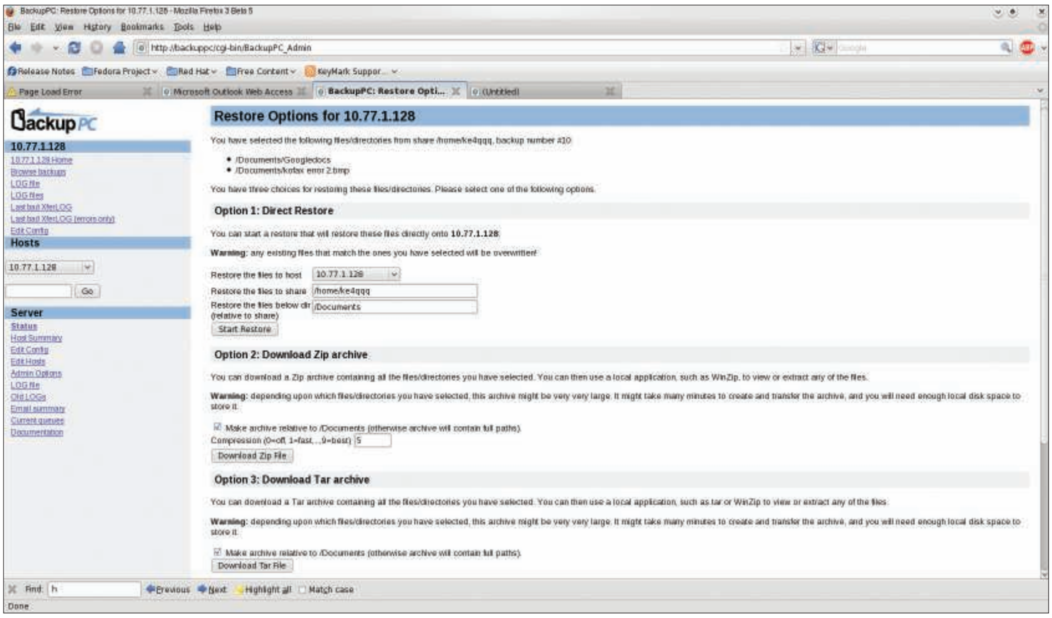
Backup browse for 10.77.1.128

- You are browsing backup #10, which started around 5/5 11:45 (21.6 days ago).
- This display is merged with backup #8.
- Select the backup you wish to view: #10 - (5/5 11:45)
- Enter directory: /Documents Go
- Click on a directory below to navigate into that directory.
- Click on a file below to restore that file.
- You can view the backup history of the current directory.

Contents of /home/ke4qqq/Documents

	Name	Type	Mode	#	Size	Date modified
□	20080108 BackupPC SAN Components v1_0.doc	file	0664	8	32768	2008-01-08 13:28:03
□	20080108 BackupPC SAN Components v1_1.doc	file	0664	8	32768	2008-01-09 13:28:42
□	Blog Writing	dir	0775	10	4096	2008-03-19 12:24:51
□	Cluster_Logical_Volume_Manager.pdf	file	0600	8	804818	2008-03-17 17:10:25
□	Googledocs	file	0664	8	1710	2008-03-19 12:59:31
□	kofax error 2.bmp	file	0664	8	1216566	2008-01-03 09:05:50
□	Landscape - Small.jpg	file	0664	8	899446	2005-03-23 10:12:28
□	My Music	dir	0775	10	4096	2007-12-28 19:23:52
□	My Pictures	dir	0775	10	4096	2008-04-08 18:26:19
□	My Videos	dir	0775	10	4096	2007-12-28 19:23:52
□	NalleysNards	dir	0775	10	4096	2008-03-18 19:06:25
□	Quote_413617253.html	file	0664	8	20567	2008-02-08 16:01:19
□	SC836.pdf	file	0600	8	16452961	2008-01-15 15:46:32
□	SCN.tif	file	0664	8	126732	2008-01-03 09:05:49
□	sds	file	0664	8	1145	2008-03-19 01:31:51
□	sdsdk	file	0664	8	241	2008-03-05 17:22:13
□	~$080108 BackupPC SAN Components v1_1.doc	file	0664	8	162	2008-01-09 13:28:42
□	Select all	Restore selected files				

Figure 5: On the Backup Browse page, you can select a single file or multiple files.



http://backuppc/cgi-bin/BackupPC_Admin

Restore Options for 10.77.1.128

You have selected the following files/directories from share /home/ke4qqq, backup number #10

- /Documents/Googledocs
- /Documents/kofax error 2.bmp

You have three choices for restoring these files/directories. Please select one of the following options.

Option 1: Direct Restore

You can start a restore that will restore these files directly onto **10.77.1.128**.

Warning: any existing files that match the ones you have selected will be overwritten!

Restore the files to host: 10.77.1.128
Restore the files to share: /home/ke4qqq
Restore the files below dir (relative to share): /Documents
Start Restore

Option 2: Download Zip archive

You can download a Zip archive containing all the files/directories you have selected. You can then use a local application, such as WinZip, to view or extract any of the files.

Warning: depending upon which files/directories you have selected, this archive might be very very large. It might take many minutes to create and transfer the archive, and you will need enough local disk space to store it.

☑ Make archive relative to /Documents (otherwise archive will contain full paths).
Compression (0=off, 1=fast,...,9=best) 5
Download Zip File

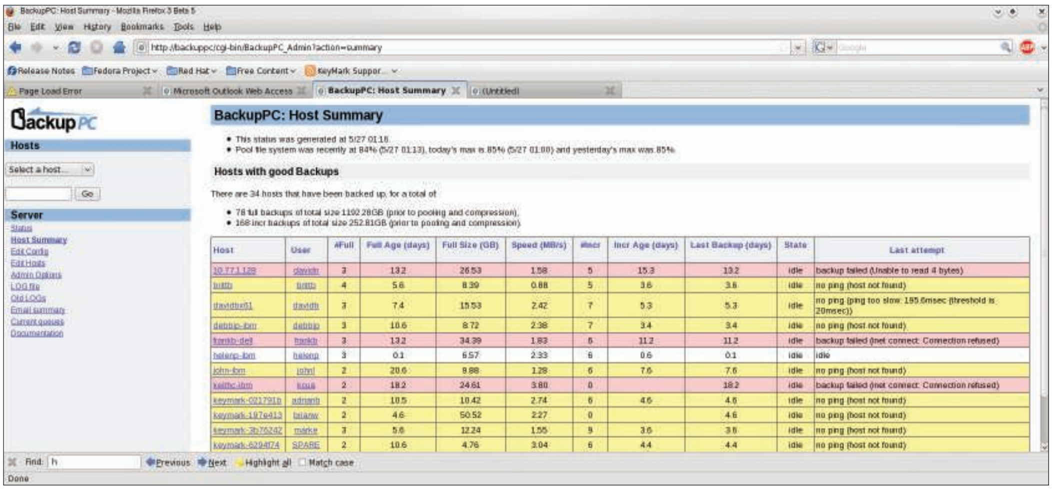
Option 3: Download Tar archive

You can download a Tar archive containing all the files/directories you have selected. You can then use a local application, such as tar or WinZip to view or extract any of the files.

Warning: depending upon which files/directories you have selected, this archive might be very very large. It might take many minutes to create and transfer the archive, and you will need enough local disk space to store it.

☑ Make archive relative to /Documents (otherwise archive will contain full paths).
Download Tar File

Figure 6: The user interface makes restoring files a breeze.



http://backuppc/cgi-bin/BackupPC_Admin?action=summary

BackupPC: Host Summary

- This status was generated at 5/27 01:18.
- Pool file system was recently at 84% (5/27 01:13), today's max is 85% (5/27 01:00) and yesterday's max was 85%.

Hosts with good Backups

There are 34 hosts that have been backed up, for a total of:

- 78 full backups of total size 1192.28GB (prior to pooling and compression),
- 168 incr backups of total size 252.81GB (prior to pooling and compression).

Host	User	#Full	Full Age (days)	Full Size (GB)	Speed (MB/s)	#Incr	Incr Age (days)	Last Backup (days)	State	Last attempt
10.77.1.128	davidn	3	13.2	26.53	1.58	5	15.3	13.2	idle	backup failed (Unable to read 4 bytes)
brittb	brittb	4	5.6	8.39	0.88	5	3.6	3.6	idle	no ping (host not found)
davidtx61	davidb	3	7.4	15.53	2.42	7	5.3	5.3	idle	no ping (ping too slow: 195.6msec (threshold is 20msec))
debbip-ibm	debbip	3	10.6	8.72	2.38	7	3.4	3.4	idle	no ping (host not found)
frankb-dell	frankb	3	13.2	34.39	1.83	6	11.2	11.2	idle	backup failed (net connect: Connection refused)
helenp-ibm	helenp	3	0.1	6.57	2.33	6	0.6	0.1	idle	idle
john-ibm	johnl	2	20.6	9.88	1.28	6	7.6	7.6	idle	no ping (host not found)
keithc-ibm	kissa	2	18.2	24.61	3.80	0		18.2	idle	backup failed (net connect: Connection refused)
keymark-021791b	adminb	2	10.5	10.42	2.74	6	4.6	4.6	idle	no ping (host not found)
keymark-197e413	brianw	2	4.6	50.52	2.27	0		4.6	idle	no ping (host not found)
keymark-3b75242	marke	3	5.6	12.24	1.55	9	3.6	3.6	idle	no ping (host not found)
keymark-6294f74	SPARE	2	10.6	4.76	3.04	6	4.4	4.4	idle	no ping (host not found)

Figure 7: A color-coded summary page gives admins an overview of backups.