

SolrCollectionDistributionScripts

This document describes the ssh/rsync based replication available since [Solr1.1](#). This mechanism only works on systems that support removing open hard links.

⚠ These Scripts were superseded by the [ReplicationHandler](#) Java implementation of index replication that works over HTTP and was introduced in [Solr1.4](#), and are no longer actively maintained. ⚠

- [snapshotter](#)
- [rsyncd-enable](#)
- [rsyncd-disable](#)
- [rsyncd-start](#)
- [rsyncd-stop](#)
- [snappuller-enable](#)
- [snappuller-disable](#)
- [snappuller](#)
- [snapinstaller](#)
- [snapcleaner](#)

All Solr collection distribution scripts and configuration files are bundled within a Solr release.

All scripts reside in the directory **`solr/bin`** of each instance of Solr. Configuration files reside in the directory **`solr/conf`** of each instance of Solr.

Snapshots directories are named **`snapshot.yyyymmddHHMMSS`**. In addition to installing a snapshot, `*snapinstaller_*` will also trigger Solr to open a new Searcher after it has put the latest snapshot into place.

A rsyncd daemon needs to be first enabled and then started on the master Solr server to handle collection distribution requests from the slaves. The file `solr/logs/rsyncd-enabled*` determines if the rsyncd daemon is enabled or disabled. The scripts `*rsyncd-enable*` and `*rsyncd-disable*` should be used to enable and disable the rsyncd daemon. The script `*rsyncd-start*` and `*rsyncd-stop*` should be used to start and stop the rsyncd daemon. Configuration of the rsyncd daemon is controlled by the file `solr/conf/rsyncd.conf*` which is generated by `*rsyncd-start*` at startup time and removed by `*rsyncd-stop*` at shutdown time. The process id of the rsyncd daemon is written into the file `solr/logs/rsyncd.pid*`. Output of the rsyncd daemon is written into the file `solr/logs/rsyncd.log*`. You may also want to install an init script on your system which starts the rsyncd daemon at system boot time if the rsyncd daemon has been enabled. Otherwise, you will have to manually start the rsyncd daemon after each system reboot.

On the slaves all the files in the index directory are hard links to the latest snapshot. This way we can keep multiple snapshots on each slave without the need to keep multiple copies of index files that have not changed. In addition, under this design the index directory will look somewhat similar on the master and slaves. On the master, all the files in the index directory will be hard links to the latest snapshot before any updates come in.

The file `solr/logs/snappuller-enabled*` determines if the `*snappuller*` is enabled or `*disabled*`. The scripts `*snappuller-enable*` and `*snappuller-disable*` should be used to enable and disable the `*snappuller*`.

The optional configuration file `solr/conf/scripts.conf*` is sourced by all scripts to set the following environment variables:

variable	description	required?	default
user	user id under which the scripts will run. If set and the invoking user id is different, the scripts will use <code>*sudo*</code> to execute itself under the specified user id. Scripts must run under an id with write permission to the data directory in order to create and install snapshots. In addition, scripts also need write permission to <code>*solr/conf*</code> and <code>*solr/logs*</code> .	optional	
solr_hostname	hostname used by scripts to connect to Solr server	optional	localhost
solr_port	port number used by scripts to connect to Solr server	required	
rsyncd_port	rsyncd port number used for distribution	optional	10000+<Solr port number>
data_dir	directory containing Lucene index used by local Solr server	optional	<code>solr/data</code>
webapp_name	webapp name used by scripts to connect to Solr server	optional	<code>solr</code>
master_host	Solr master server hostname	required for distribution	
master_data_dir	directory containing Lucene index on the master Solr server	required for distribution	
master_status_dir	directory containing distribution related status on the master Solr server	required for distribution	

Required variables must be set either in `*solr/conf/scripts.conf*` or by command line arguments. The values set by command line arguments will override default values and values set in `*solr/conf/scripts.conf*`.

snapshotter

```
usage: snapshotter [-d dir] [-u username] [-v]
  -d    specify directory holding index data
  -u    specify user to sudo to before running script
  -v    increase verbosity
  -V    output debugging info
```

*

snapshotter* takes a snapshot of the current Lucene index by using hard links. Snapshots directories are named* snapshot.yyyymmddHHMMSS* where_* yyyymmddHHMMSS_* is the timestamp of when the snapshot was taken. All relative paths specified with the -d option start at the Solr root directory *solr_'.

rsyncd-enable

```
usage: rsyncd-enable [-u username] [-v]
  -u    specify user to sudo to before running script
  -v    increase verbosity
  -V    output debugging info
```

rsyncd-enable enables the starting of the rsyncd daemon by creating the file **solr/logs/rsyncd-enabled**. Please note that this script will not actually starts the rsyncd daemon.

rsyncd-disable

```
usage: rsyncd-disable [-u username] [-v]
  -u    specify user to sudo to before running script
  -v    increase verbosity
  -V    output debugging info
```

rsyncd-disable disables the starting of the rsyncd daemon by removing the file **solr/logs/rsyncd-enabled**. Please note that this script will not actually stop the rsyncd daemon if it is already running.

rsyncd-start

```
usage: rsyncd-start [-d dir] [-p portnum] [-u username] [-v]
  -d    specify directory holding index data
  -p    specify rsyncd port number
  -u    specify user to sudo to before running script
  -v    increase verbosity
  -V    output debugging info
```

Starts the rsyncd daemon on the master Solr server. The rsyncd daemon sets its port number to be the port number of the Solr server incremented by 10000 if the environment variable **rsyncd_port** is not set in **solr/conf/scripts/conf** and the **-p** command line argument is not specified. For example, if the Solr server runs at port 7000, then its rsyncd daemon runs at port 17000. The start script is synchronous. After starting the rsyncd daemon, it will attempt to connect to it for up to 15 seconds. The start script will exit with error code 2 if it fails to connect to the rsyncd daemon.

Before starting the rsyncd daemon, this script generates the configuration file **solr/conf/rsyncd.conf** which is in turn removed by **rsyncd-stop** at shutdown time. The **rsyncd.conf** specifies a module that exports a directory tree via the daemon, which limits rsync clients to the **solr.home/data directory**. A side effect is that rsync clients connect to a directory named **solr**, which is a virtual path. See the [rsyncd.conf](#) man page for more information on the format of this file.

The process ID of the rsyncd daemon is written into the file **solr/logs/rsyncd.pid**. Output of the rsyncd daemon is written into the file **solr/logs/rsyncd.log**. All relative paths specified with the -d option start at the Solr root directory **solr**.

rsyncd-stop

```
usage: rsyncd-stop [-u username] [-v]
  -u    specify user to sudo to before running script
  -v    increase verbosity
  -V    output debugging info
```

Stops the rsyncd daemon on the master Solr server. The stop script is synchronous. After stopping the rsyncd daemon, it makes sure that the daemon has exited by trying to connect to it for up to 300 seconds. The stop script will exit with error code 2 if it fails to stop the rsyncd daemon. After the rsyncd daemon has been stopped, it removes the configuration file **solr/conf/rsyncd.conf** which is generated by **rsyncd-start** at startup time.

snappuller-enable

```
usage: snappuller-enable [-u username] [-v]
      -u      specify user to sudo to before running script
      -v      increase verbosity
      -V      output debugging info
```

snappuller_enable enables the snappuller by creating the file **solr/logs/snappuller-enabled**.

snappuller-disable

```
usage: snappuller-disable [-u username] [-v]
      -u      specify user to sudo to before running script
      -v      increase verbosity
      -V      output debugging info
```

snappuller-disable disables the snappuller by removing the file **solr/logs/snappuller-enabled**.

snappuller

```
usage: snappuller [-M master] [-P portnum] [-D mdir] [-S sdir] [-n snapshot] [-d dir] [-u username] [-svz]
      -M master    specify hostname of master server from
                   where to pull index snapshot
      -P port      specify rsyncd port number of master
                   server from where to pull index
                   snapshot
      -D           specify directory holding index data
                   on master server
      -S           specify directory holding snapshot
                   status on master server
      -n snapshot  pull a specific snapshot by name
      -d           specify directory holding index data
                   on local machine
      -u           specify user to sudo to before running
                   script
      -s           use the --size-only option with rsync
      -v           increase verbosity (-vv show file
                   transfer stats also)
      -V           output debugging info
      -z           enable compression of data
```

If **snappuller** has been disabled, it will log an appropriate message in its log file, and then exit without pulling any snapshot from the master Solr server.

If the name of the snapshot to be pull is not specified by the use of the "-n" option, **snappuller** will use ssh to determine the name of the most recent snapshot available on the master Solr server and pull it over if it does not already exist on the slave Solr server.

The status and stats of the current or most recent rsync operation of snappuller is kept in the file **solr/logs/snappuller.status**. Whenever this file is updated by snappuller, a copy is scp back to the master Solr server. See [SolrCollectionDistributionStatusStats](#) for more details.

All relative paths specified with the -D and -d options start at the Solr root directory **solr**.

snapinstaller

```
usage: snapinstaller [-M master] [-S sdir] [-d dir] [-u username] [-v]
  -M master  specify hostname of master server from
              where to pull index snapshot
  -S         specify directory holding snapshot
              status on master server
  -d         specify directory holding index data on
              local machine
  -u         specify user to sudo to before running
              script
  -v         increase verbosity
  -V         output debugging info
```

After a snapshot has been installed, **snapinstaller** writes its name into the file **`solr/logs/snapshot.current`**, and scp a copy of this file back to the master Solr server. See [SolrCollectionDistributionStatusStats](#) for more details. All relative paths specified with the `-d` option start at the Solr root directory **`solr`**.

snapcleaner

```
usage: snapcleaner -D <days> | -N <num> [-d dir] [-u username] [-v]
  -D <days> cleanup snapshots more than <days> days
              old
  -N <num>   keep the most most recent <num> number
              of snapshots and cleanup up the
              remaining ones that are not being pulled
  -d         specify directory holding index data
  -u         specify user to sudo to before running
              script
  -v         increase verbosity
  -V         output debugging info
```

Please note that the argument to the command line option `-D` is not inclusive. In addition, a day old is defined as more than 24 hours old from the time the command is invoked. For example, `-D 1` will cause **snapcleaner** to remove snapshots more than but not including 1 day old, ie 2 days and older. Therefore, snapshots more than 48 hours old will be removed. All relative paths specified with the `-d` option start at the Solr root directory **`solr`**.