
Using rsync to Maintain a Warm DataFax Backup

Phil Kirsch
DF/Net Research, Inc.



In case of a real disaster it takes time to ...

- Purchase equipment
- Copy data from backup
- Perform cursory checks of data integrity

Few would even think about
validation

Until the next audit

There is a better way

- Secure, remote infrastructure is available online
- rsync can compliantly update a remote DataFax system
- ssh provides security
- DataFax Reports can verify equivalence
- Perl ties it together
- Cron provides automation
- Sendmail provides notification

'Cloud' Infrastructure (Utility Computing)

- Available from scores of companies
 - Up to 450 different operating systems
 - For around \$150 to \$500 per month
- Or use a colo to create your own
- As secure as you make it
 - tcp 'wrappers'
 - Firewalls
 - ssh
 - Etc.
 - Certifications, audits, etc.

rsync Features

- Security (*ssh*)
- Speed (*only transfers changed bytes*)
- Logging (*documentation*)
- Specific file names or patterns can be omitted (e.g. "*DFserver.cf*" or "**.tar*")

DataFax Reports

- Run the following reports just prior to transfer and write to file (*minus time and server name*).
 - DFstatus counts records by status and level
 - DF_ICimages compares data with CRFs
 - DF_ICqcs counts queries
 - DF_ICrecords compares data to setup
- Copy report to the remote server
- Run the same reports on remote server
- Use 'diff' to compare the output.

Perl controls step-wise execution

1. Checks .serverstatus## for study activity
2. Creates dated log and report file names
3. Runs DataFax reports on production
4. Removes unique elements before writing
5. Stops datafax on remote system via ssh
6. Calls rsync with designated options
7. Restarts everything on both systems
8. Runs DataFax reports on remote system
9. Compares production output to remote
10. Emails notification of results

Sample rsync log output

```
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-163212
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-163313
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-163413
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-163534
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-164418
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-164521
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-171758 ← New files
2011/11/16 00:08:10 [1685] <+++++ 035/dfschema/111115-171956
2011/11/16 00:08:11 [1685] <+++++ 035/dfschema/111115-172420
2011/11/16 00:08:11 [1685] <+++++ 035/dfschema/111115-173428
2011/11/16 00:08:11 [1685] <+++++ 035/dfschema/111115-173530
2011/11/16 00:08:11 [1685] <+++++ 035/dfschema/111115-185306
2011/11/16 00:08:11 [1685] <+++++ 035/dfschema/111115-185700
2011/11/16 00:08:11 [1685] <+++++ 035/dfschema/111115-195203
2011/11/16 00:08:11 [1685] .d.t..... 035/df/
2011/11/16 00:08:11 [1685] <f.t..... 035/df/DFUnexpected.drf
2011/11/16 00:08:11 [1685] <f.t..... 035/df/DFIllegal.drf
2011/11/16 00:08:11 [1685] <f.t..... 035/df/DFincon.drf
2011/11/16 00:08:11 [1685] <f.st..... 035/ecsrc/DFedits
2011/11/16 00:08:11 [1685] .d.t..... 035/lib/
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/DFedits.bin
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/DFfile_map
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/DFschema
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/DFschema.st ← Updated Files
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/DFsetup
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/DFtips
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/batch_edits_test.xml ← Edit check output readable only by programmer
2011/11/16 00:08:11 [1685] <f.st..... 035/lib/batch_edits_test_output.html
2011/11/16 00:08:11 [1685] rsync: send_files failed to open "/opt/studies/035/lib/batch_edits_test_output.xml": Permission denied (13)
2011/11/16 00:08:11 [1685] .d.t..... 035/pages/0842/
2011/11/16 00:08:11 [1685] *deleting 035/pages/0842/0060002
2011/11/16 00:08:11 [1685] <+++++ 035/pages/0842/X0060002
2011/11/16 00:08:11 [1685] .d.t..... 035/pages/0906/
2011/11/16 00:08:11 [1685] *deleting 035/pages/0906/005L001
2011/11/16 00:08:11 [1685] <+++++ 035/pages/0906/X005L001 ← Deleted CRF
2011/11/16 00:08:11 [1685] .d.t..... 035/pages/0909/
2011/11/16 00:08:11 [1685] *deleting 035/pages/0909/00LB001
2011/11/16 00:08:11 [1685] <+++++ 035/pages/0909/X00LB001
```

Administrative Requirements

- Each study needs to be defined twice in DFsystem.
- DataFax permissions are not transferred and require definition before system use.
- Functions called by edit checks will also need to be transferred if they do not exist within the remote system's OS.
- Edit check output could require modification depending on configuration of the remote system.

The End Result is a Remote System

- Where the transfer of each CRF image, data and journal file is logged.
- Where creation, modification and deletion of files is logged.
- Where the congruence of images and records is verified and documented daily.
- Where the number of records and QCs at each level and status are documented to match production at a specific time point.

In other words ..

there is documented evidence that
it exactly matches the production
system at a known time point.