Migrating DataFax from a 32-bit server to a 64-bit server

Michael Holdsworth
All NIAID Linux servers are managed by one team and they required us to migrate to a 64-bit server.

• In November 2015 we started a project to migrate DataFax from a 32-bit server (running Red Hat version 5.1) to a 64-bit server (running Red Hat version 6.6).
What is 32-bit and 64-bit?

• To keep it simple, the whole bit thing (16-bit, 32-bit, 64-bit) refers to how much data the computer can keep track of, or talk to, at once.

• And this is what determines how much memory (RAM - Random Access Memory) it can handle.

• A processor with 32-bit memory can deal with 4GB of RAM.

• A 64-bit system, on the other hand, can rock 16 exabytes of RAM.

• That's 16.8 million terabytes of RAM.

• With 64-bit hardware and software, you can use vast amounts of RAM, which enables a whole new world of possibilities for applications:
  ◦ Improved performance
  ◦ Improved capabilities
  ◦ Dramatically increased memory
  ◦ Power efficient
  ◦ Improved security
Reasons for Migrating to 64-bit

• DataFax is a 32-bit software, so there is no immediate benefit (from a DataFax perspective) gained from migrating to a 64-bit server.

• However, from a server administration perspective, there are good reasons for migrating:
  • Server Efficiencies
  • Server Capabilities
  • Operational Logistics
  • Server Security
Approach to the Project

We approached the project by:

- Initially setting up a 64-bit virtual server (running Red Hat 6.6) in a development environment.
- Installing DataFax 2014.1.0 onto the new 64-bit server.
- Validating the DataFax software installation by completing the full ATK (DF/Net’s Acceptance Test Kit).
Winston Churchill

When asked to define history...

History is simply one damned thing after another.

- Winston Churchill
Lessons Learnt

1. DataFax related software components
2. Server level folders/files to be synchronized
3. Detailed Testing protocol to validate synchronization
4. Server level synchronization commands
5. Changing DFstudies.db to point to “localhost”
6. Synchronizing the archive directory
7. DataFax related processes
Lesson 1: DataFax related software components

• Initially, we setup a new 64-bit server and installed DataFax 2014.1.0.

• However, DataFax software does depend on other software components (like Java).

• We discovered, as we were working through the ATK, that we had installed Java in the wrong directory.

• This negatively effected some of DataFax’s functionality.
Lesson 2: Server level folders/files to be synchronized

1. `opt/studies` (granular study related data)
2. `/opt/datafax/lib/DFstudies.db` (study definitions)
3. `/opt/datafax/lib/DFuserdb.log` (user database log)
4. `/opt/datafax/lib/DFuserdb.idx` (user database index)
5. `/opt/datafax/identify` (unidentified fax router)
6. `/opt/datafax/incoming` (incoming new faxes – synchronize if this directory is not empty)
7. `/opt/datafax/work/(all .seq)` (sequence files)
8. `/opt/datafax/work/fax_log` (fax log)
9. `/opt/datafax/work/fax_log.idx` (fax log index)
10. `/opt/datafax/archive`
Lesson 3: Detailed Testing protocol to validate synchronization

• We had to develop a detailed testing protocol to verify that all data/metadata had been successfully synchronized from our 32-bit server to our new 64-bit server.

• We had tried to synchronize the two servers many times, so every time we ran a synchronization we had to setup new **test cases**.

• We also had to synchronize when no users were in the system.

• Our testing protocol had to ensure that the following aspects had been successfully synchronized:
  a. All study data/metadata
  b. Unidentified fax router
  c. All audit trail information
  d. All users
  e. User permissions
  f. Study definitions

After every synchronization, we would also complete a miniATK (a smaller version of the full ATK) to ensure that DataFax critical functionality was still working.
Lesson 4: Server level synchronization commands

- To synchronize the 2 servers we, mostly, used a command line tool called rsync.
- Rsync is a fast and extraordinarily versatile file copying tool.
- It offers a large number of options that control every aspect of its behavior and permit very flexible specification of the set of files to be copied.
- However, because rsync is so powerful and flexible it took us a few failed synchronizations to get the arguments correct.
- Initially, we had problems with:
  a. Files not being overwritten
  b. Permissions not being preserved
  c. Groups/ownerships not being preserved
Lesson 5: Changing DFstudies.db to point to “localhost”

- DFstudies.db is a list of all current studies, with related study configuration information.
- $DATAFAX_DIR/lib/DFstudies.db
Lesson 6: Synchronizing the archive directory

- Initially, we did not synchronize the archive directory (that is /opt/datafax/archive).
- Theoretically, it should not make a difference.
- However, there are situations when it can cause problems:
  
  - We were working through module 10 of the ATK and pushed the module button (in the ATK tool) called "Fax QC Corrections" which is supposed to submit 5 pages to the ATK study.
  - However, we discovered that the pages never reached the ATK study.
  - Instead, they were stuck in the /opt/datafax/work/001 directory.
  - The DFincoming process was trying to archive the images in a file called /opt/datafax/archive/1049/001H, which already existed.
  - So, the process was bombing out and logging an error message in /var/log/messages.
Lesson 7: DataFax related processes

• Every DataFax installation has various related processes, depending on the environment.

• In the case of our NIAD installation, we have CronJobs or Msource feeds.

• Migrating DataFax to a new server architecture or a new OS (operating system) version can effect these processes.

• For example:
  • Many of our CronJobs (and CronJob referenced scripts) used an emailing tool called “mailx”.
  • Somewhere in between Red Hat version 5.1 and Red Hat version 6.6 the “mailx” tool had been updated so that all referenced email addresses have to be preceded with a double dash “--”.
DF/NET SUPPORT TEAM

We would like to acknowledge DF/Net Research, and specifically Craig Dilks, for their ongoing and indispensable support during the project.

NIAID TEAMS

IBRSP – International Biomedical Research Support Program

OEB Linux - Operations & Engineering Branch

OCICB - Office of Cyber Infrastructure and Computational Biology
Questions

Winston Churchill

A good speech should be like a woman's skirt: long enough to cover the subject and short enough to create interest.