

Instructions: Upgrade to latest Fusion 2.1

This set of upgrade instructions is for

- FUSION-CURRENT: Fusion versions 1.2.3 through 2.0.0
- FUSION-NEW: the latest version of Fusion 2.1

INSTALL-DIR refers to the parent directory of the FUSION-CURRENT deployment or a staging directory which contains a copy of the FUSION-CURRENT installation. All commands in the upgrade instruction set are run from this directory.

FUSION-UPGRADE-SCRIPTS is the full path to the download for this repo. The conversion scripts and programs are in directory FUSION-UPGRADE-SCRIPTS/bin.

Upgrade Process Overview

Preparation

- [Download, unpack FUSION-NEW](#)
- [Copy embedded ZooKeeper and/or Solr](#) If Fusion is using either the ZooKeeper or the Solr installation included with the Fusion distribution or both, the ZooKeeper and Solr data must be copied from FUSION-CURRENT to FUSION-NEW.

Reconfiguration

- [Save custom settings in Fusion configuration files and run scripts](#). Preserve customizations, including settings for external ZooKeeper and Solr.
- [Copy JDBC driver jarfiles](#). Only required for JDBC datasources.
- [Propagate datasource and pipeline configurations](#). The script [download_upload_ds_pipelines.py](#) transfers this information between the current and new ZooKeeper.
- [Copy and convert the crawldb](#). The crawldb records indexing job runs, and this job history must be carried forward and the data format must be converted to the format used in Fusion 2.1 via the conversion utility [com.lucidworks.fusion-crawldb-migrator-0.1.0.jar](#).

Preparation

To upgrade while leaving FUSION-CURRENT deployment in place, you must set up parallel current and new Fusion installs in the INSTALL-DIR. If either the local ZooKeeper or Solr installation is being used, this data must be copied over from the current to the new installation.

Download, unpack FUSION-NEW

`cd` to directory `INSTALL-DIR`. The disk partition this directory is on must have at least as much free disk space as the size of the `FUSION-CURRENT` directory. On a *nix system, the following commands can be used:

- `du -sh fusion` - total size of `FUSION-CURRENT`.
- `df -kH` - amount of free space on all file-systems.

Download or copy the Fusion distribution for `FUSION-NEW` into `INSTALL-DIR`. The latest Fusion distribution is available from <https://lucidworks.com/products/fusion/download/get-started/> It is distributed as a gzipped tar file or as a compressed zip file.

Create a new directory named "fusion-new" and unpack the contents of the distribution here. For example, to upgrade to Fusion 2.1.1, the archive file is named "fusion-2.1.1.tar.gz". The default `tar -xf` command would unpack this into a directory named "fusion" which would overwrite parts of the `FUSION-CURRENT` installation. To avoid this, run the following commands:

```
> mkdir fusion-new
> tar -C fusion-new --strip-components=1 -xf fusion-2.1.1.tar.gz
```

If you are working on a Windows machine, the zipfile unzips into a folder named "fusion-2.1.1" which contains a folder named "fusion". Rename "fusion" to "fusion-new" and moved it into folder `INSTALL-DIR`.

Copy embedded ZooKeeper and/or Solr (as needed)

The Fusion distribution includes both a ZooKeeper and Solr install. This "out of the box" configuration ("OOTB") is intended mainly for development purposes. In a production an external ZooKeeper cluster should be used, and an external Solr cluster is likely to be used.

If you are running the OOTB Fusion, then you must copy over both the ZooKeeper configuration and all Solr data. If you are using an external ZooKeeper but using Fusion's local Solr install, you must copy over all Solr data.

To copy the ZooKeeper configuration:

```
> mkdir -p fusion-new/data/zookeeper
> cp -R fusion/solr/zoo_data/* fusion-new/data/zookeeper
```

To check your work: compare the directories `fusion/solr/zoo_data/` and `fusion-new/data/zookeeper` using the `diff` command. This command succeeds silently when the contents are the same.

```
> diff -r fusion/solr/zoo_data fusion-new/data/zookeeper
```

To copy the Solr data:

```
> find fusion/solr -maxdepth 1 -mindepth 1 | grep -v -E "zoo*" | while read f ; do cp -R $f fusion-new/data/solr/; done
```

If the Solr collections are very large this may take a while.

You can use the `diff` command to check your work. The copy command excluded ZooKeeper config data, therefore you should see the following output:

```
> diff -r fusion/solr fusion-new/data/solr
Only in fusion-new/data/solr: configsets
Only in fusion/solr: zoo.cfg
Only in fusion/solr: zoo_data
```

Reconfiguration

Step 1. Save custom settings in Fusion configuration files and run scripts

The Fusion configuration files define environment variables used in the Fusion run scripts. The Fusion run scripts start and stop Fusion services. If you are using an external ZooKeeper, or have changed memory limits on any Fusion services, these files will have been changed.

In Fusion 1.2, the "fusion/bin" directory contains configuration files and run scripts for all Fusion services:

```
README.txt
api
api.cmd
common.sh
config.cmd
config.sh
connectors
connectors.cmd
fusion
fusion.cmd
oom.sh
solr
solr.cmd
spark-master
spark-master.cmd
spark-worker
spark-worker.cmd
ui
ui.cmd
```

If these files have been customized for your installation, you must identify those changes and then **edit** the files in "fusion-new" accordingly. New changes and optimizations have been added to these files for the 2.1 release. These changes will be lost by replacing the new with the old by a copy command, thus a by-hand edit is needed to properly carry over customizations.

To facilitate the task of identifying changes made to the current installation, the fusion-upgrade-scripts repository contains a directory "reference-files" which contains bin directories for Fusion releases 1.2.3, 1.2.4, and 1.2.6 named "bin-1.2.3", "bin-1.2.4", and "bin-1.2.6". To identify changes, use the *nix **diff** command with the **-r** flag, e.g. if FUSION-CURRENT is 1.2.3, then the command is:

```
> diff -r INSTALL-DIR/fusion/bin FUSION-UPGRADE-SCRIPTS/reference-files/bin-1.2.3
```

In Fusion 2.1, the configuration files "config.sh" and "config.cmd" have been moved to directory "fusion-new/conf" and the scripts to run the Fusion services remain in "fusion-new/bin".

NOTE

if you are running external ZooKeeper (recommended for production system), you should edit the Solr start script in file "fusion-new/bin/solr" and delete the command which starts the embedded ZooKeeper. This line is: ``-DzkRun \ ``

Step 2. Copy JDBC driver jarfiles

The jarfiles for any JDBC drivers used by a JDBC datasource are found in directory: "fusion/data/connectors/lucid.jdbc" Copy the contents of this directory over to the "fusion-new" directory:

```
> cp -R fusion/data/connectors/lucid.jdbc fusion-new/data/connectors/
```

Step 3. Propagate datasource and pipeline configurations

The script [download_upload_ds_pipelines.py](#) is used to propagate Fusion datasource and pipeline configurations stored in ZooKeeper, in two stages:

1. download the configuration information from ZooKeeper to local directory named "fusion_upgrade_2.1".
2. upload the configuration information to the Fusion API service.

This script uses the following arguments and values:

- "--zk-connect": the ZooKeeper server:port for FUSION-NEW
- "--action": either "download" or "upload".
- "--fusion-url": URL of Fusion API service to upload configurations to
- "--fusion-username": name of Fusion user with admin privileges
- "--fusion-password": password for Fusion user

Download configurations from ZooKeeper

No services for FUSION-NEW should be running, except for ZooKeeper. If your Fusion installation uses an external ZooKeeper, then this must be running. If your Fusion installation uses an embedded ZooKeeper, then you must have copied the ZooKeeper data from FUSION-CURRENT to FUSION-NEW (see [link:#prep-2](#)[instructions above]).

Start the ZooKeeper service:

```
> fusion-new/bin/zookeeper start
```

Run the script to download the configurations.

```
> python FUSION-UPGRADE-SCRIPTS/bin/download_upload_ds_pipelines.py \  
--zk-connect localhost:9983 --action download
```

To check your work, check that directory "fusion_upgrade_2.1" was created and that it contains definitions for all datasources and pipelines. Do not remove this directory until you have successfully completed the upload step.

If you are running embedded ZooKeeper, shut it down again:

```
> fusion-new/bin/zookeeper stop
```

Upload configurations to the Fusion API service

Start FUSION-NEW:

```
> fusion-new/bin/fusion start
```

Once it is running, run the script in upload mode to propagate the configurations in directory "fusion_upgrade_2.1".

Because the Fusion usernames and password are stored in ZooKeeper, the FUSION-NEW will have the same admin user name and password. To upload data to the Fusion API services, you must supply these as arguments to the script:

- "--fusion-username": name of Fusion user with admin privileges
- "--fusion-password": password for Fusion user

```
> fusion-new/bin/fusion start
> python FUSION-UPGRADE-SCRIPTS/bin/download_upload_ds_pipelines.py \
  --zk-connect localhost:9983 --action upload --fusion-url http://localhost:8764/api \
  --fusion-username <admin> --fusion-password <pass>
```

Step 4. Copy and convert the crawldb

The Fusion "crawldb" records the results of running datasource jobs. This information must be copied from FUSION-CURRENT to FUSION-NEW.

Copy the Fusion "crawldb" directory:

```
> cp -R fusion/data/connectors/crawldb fusion-new/data/connectors/
```

The crawldb data format changed in Fusion 2.1, therefore to upgrade to 2.1, the crawldb data must be converted with the the conversion utility [com.lucidworks.fusion-crawldb-migrator-0.1.0.jar](https://github.com/lucidworks/fusion-crawldb-migrator).

The **anda-v1-to-v2** command allows Fusion 1.2.x connector DBs to be updated to the new v2.x format. It requires:

- A Fusion pre 2.1 install ("the v1 install", here, FUSION-CURRENT)
- A Fusion 2.1 install ("the v2 install", here, FUSION-NEW). The Fusion 2.1 install must have:
 - All v1 Datasources must have been propagated to the new v2 instance

- All v1 crawldb files must have been copied over to the new v2 instance

The Fusion 2.1 install must be running. If the FUSION-NEW installation is not currently running, start it:

```
> fusion-new/bin/fusion start
```

The **anda-v1-to-v2** takes the following arguments:

- path-to-FUSION-CURRENT
- path-to-FUSION-NEW
- the -z flag specifies the ZooKeeper server:port for FUSION-NEW

The command to run this utility from the INSTALL-DIR is:

```
> java -jar FUSION-UPGRADE-SCRIPTS/bin/com.lucidworks.fusion-crawldb-migrator-0.1.0.jar  
anda-v1-v2 fusion fusion-new -z localhost:9983
```

Once the task successfully completes, **the last few lines of logging show the output directory of the new DB files**. The output must be copied over to FUSION-NEW. To do this, remove the existing **lucid.anda** db directories, then copy the new **lucid.anda** directories generated from this utility into that same location:

```
> rm -Rf fusion-new/data/connectors/crawldb/lucid.anda/*  
> mv ${path-printed-from-command-output} fusion-new/data/connectors/crawldb/lucid.anda/
```

This completes the upgrade process.

At this point, you should validate the FUSION-NEW, per instructions in the [link:README.asciidoc](#). Once validated, you can archive and/or delete the directory INSTALL-DIR/fusion and rename INSTALL-DIR/fusion-new to INSTALL-DIR/fusion.