NDEx Sync 2.2: A Network Copier Utility

Last updated: October 20th, 2017

Overview

NDEx Sync is a command line utility that enables users to copy networks from one NDEx account (the source NDEx) to another (the target NDEx).
NDEx Sync 2.2 is fully compatible with the latest NDEx 2.2 release and no changes are required to use it.
NDEx Sync 2.2 can be downloaded from our FTP Server (ftp://ftp.ndexbio.org/NDEx-v2.2/).

Requirements

1. Platform: Linux or Mac OS X
2. Java 8 installed
3. Network access to both the source and target NDEx servers running v2.0 or higher.

Limitations

NDEx Sync 2.2 can ONLY copy networks between NDEx servers running NDEx v2.0 or higher.

Important Changes

The "route" parameter in the copy plans should use the new v2 endpoints! See the Copy Plans section below for more details.

License and Source Code

NDEx Sync is open-source software available under a BSD license. The source code is hosted on GitHub at https://github.com/ndexbio/ndex-sync (https://github.com/ndexbio/ndex-sync)

Running the NDEx Sync

NDEx Sync is used via the shell script /opt/ndex/lib/ndex-copier.sh

The script takes a single argument: a directory containing ‘copy plan' files.

```
bash ndex-copier.sh .
```

When run, the script reads and attempts to execute each copy plan file in the directory.

The NDEx Sync script can be run manually or can be executed periodically via cron or other scheduling facilities to copy new or modified networks from the source NDEx, creating or updating networks on the target NDEx.

How NDEx Sync Works

NDEx Sync is like a file-mirroring utility, but with an important difference: the copied networks are not exact duplicates of the source networks.

Copied networks are assigned new UUIDs: every network stored in an NDEx server has a globally unique identifier and can be referenced by that identifier at its host NDEx.

NDEx Sync updates (or creates, if necessary) the network's provenance history, adding a "provenance event" that documents the fact of the copying.

The copied networks are therefore documented as distinct entities, copied at a specific time from a uniquely identified source. The provenance history provides a structure to document the events leading to the current state of a network. Applications using NDEx are not required to maintain the provenance history for networks that they manipulate, but it is encouraged as a standard practice and will be supported by NDEx utilities.

For each source network that is selected as a candidate for copying, NDEx Sync examines the provenance history of each network in the target account to determine:

Was this target network copied from the source network?

Is the target Out-Of-Date?

The default behavior of NDEx Sync is that it will copy the source network to the target account if there is no copy of the source network in the target account OR if the only copies are Out-Of-Date or have been modified.

Update of Networks by NDEx Sync

The default behavior of NDEx Sync is conservative, never overwriting or deleting any network in the target directory. This behavior can be overridden by the copy plan parameter updateTargetNetwork, specifying that NDEx Sync should update target networks that are identified as unmodified, out-of-date copies of the specified source networks.

In an update, the target network keeps its UUID but its contents are replaced by the contents of the source network and the provenance history is handled in the same manner as in a default, non-update copy event. The updated network may be accessed by that UUID and any new request will obtain the updated content.

Using NDEx Sync to update networks is only appropriate for situations in which the target network is intended as a cache of the source, where users want to obtain the latest version of the source content and where they do not expect the content of the network to be consistent over time.

Updates of Read Only Networks

By default, updates will NOT be performed if the target network has readOnly == true. The updateReadOnlyNetwork configuration parameter in a copy plan overrides this behavior. This handles the case in which NDEx Sync is used to maintain a local copy of a remote resource and where the local copy is intended as a read-only reference.

Out-Of-Date Criteria

The criteria for "out-of-date" are as follows:

Calculate latestSourceDate as the later of modification date and the last provenance history event end date for the source network.

Calculate earliestTargetDate as the earlier of modification date and the last provenance history event end date for the target network.

if latestSourceDate > earliestTargetDate, target is out-of-date

Last Modification Date

The lastModificationDate field of a network is updated when: 1) there is a change to any network element, including properties and presentation properties or 2) there is a change to intrinsic special "profile" properties (name, description)

The lastModificationDate does not update on:

Changes to provenance history, changes to permissions, changes to read-only status and changes to visibility.

What is Copied with a Network

All network elements, including properties, presentation properties are copied. In addition, the Provenance History is also copied and modified.

The following elements are not copied: permissions, visibility, UUID, modification time, creation time and readOnly status.

Copy Plans

NDEx Sync ‘copy plans' specify:

An **account and credentials** for the source NDEx.

An **account and credential**s for the target NDEx.

The **criteria to select networks** on the source NDEx, which can be one of: *1) a query to find networks matching search text, 2) a query to find networks administered by an account AND matching search text or 3) a list of network UUIDs.*

The **updateTargetNetwork parameter**

- The possible values of updateTargetNetwork argument are "true" or "false".
- The default value of this argument (i.e., if is missing from the copy plan) is "false".
- If updateTargetNetwork is set to "true", NDEx Sync should check whether the target server account specified in the copy plan contains a network that was copied earlier from the source server, and decide whether to update the network in the target server account or not. In case the network only exists in the source server account and not in the target account, the network gets copied to the target account.

The **updateReadOnlyNetwork parameter**

- The value of updateReadOnlyNetwork argument is "true" or "false".
- Default value (if the argument is missing from the copy plan) is "false".
- If updateReadOnlyNetwork is true and the target account specified in the copy plan has the Administrator privileges for the target network to be updated then the target network can be updated even if it is set to readOnly = true. In this case, NDEx Sync changes the read-only flag to false, updates the network, and changes the read-only flag back to true.
- The updateReadOnlyNetwork parameter is only used if updateTargetNetwork is set to true.

Notes on Updates

NDEx Sync can only update networks in the target server account if the account specified by the username in the target element in the copy plan must have Administration privileges for the networks to be updated.

Query Copy Plan

Source networks are identified based on their title, description, or content matching a query string. The user account for the source must have read access to each source network.

In the example copy plan below, networks matching "cal*" are copied from the public NDEx to the user2 account on an NDEx running on the local machine.

queryString: search text to find networks.

queryLimit: a maximum number of networks to copy is specified. This is useful largely as a brake on runaway copying – if the queryString matched some unanticipated, enormous number of networks, the script would still be limited.

```
{
"planType" : "QueryCopyPlan",
"source" : {
"route" : "http://www.ndexbio.org/v2",
"username" : "user1",
"password" : "pwd00123"
},
"target" : {
"route" : "http://myPrivateNDExServer.com/v2",
"username" : "user2",
"password" : "pwd980098"
},
"queryString" : "cal*",
"queryLimit" : "10",
"updateTargetNetwork" : "false",
"updateReadOnlyNetwork" : "false"
}
```

Query Copy Plan with Account

sourceAccount: Source networks are limited to those administered by the specified account name.

To copy all the networks for a given account, the **queryString** can be "*"

In the copy plan example below, all networks (up to 10) from the user3 account are copied from the public NDEx to the user2 account on an NDEx running on the local machine.

```
{
"planType" : "QueryCopyPlan",
"source" : {
"route" : "http://www.ndexbio.org/v2",
"username" : "user1",
"password" : "pwd00123"
},
"target" : {
"route" : "http://myPrivateNDExServer.com/v2",
"username" : "user2",
"password" : "pwd980098"
},
"queryString" : "*",
"queryLimit" : "10",
"queryAccountName" : "user3",
"updateTargetNetwork" : "false",
"updateReadOnlyNetwork" : "false"
}
```

Network ID Copy Plan

idList: list of UUIDs to identify source networks.

The **user account** for the source must have read access to each source network.

In this example, the network 5bca3218-28ca-11e4-9032-90b11c72aefa is copied from the public NDEx to the user2 account on an NDEx running on the local machine.

```
{
"planType" : "IdCopyPlan",
"source" : {
"route" : "http://www.ndexbio.org/v2",
"username" : "user1",
"password" : "pwd00123"
},
"target" : {
"route" : "http://myPrivateNDExServer.com/v2",
"username" : "user2",
"password" : "pwd980098
},
"idList" : [
"5bca3218-28ca-11e4-9032-90b11c72aefa"
],
"updateTargetNetwork" : "false",
"updateReadOnlyNetwork" : "false"
}
```