HDFS Archive (HAR) Compactor How to Configure

On this page:

Step 1. Launch Aspire and Open the Services Management Page

1. Launch Aspire (if it's not already running).
2. Go to Launch Control.
3. Browse to: http://localhost:50505
4. Select "Services" from the drop down on the right to display the "Service Management" page

? Unknown Attachment

For details on using the **Aspire Service Management** page, see Admin UI.

Step 2. Add a New Service

1. From **Service Management**, click **Add Service** to add a new service.
    a. Or add a custom one with the group id "com.searchtechnologies.aspire" and artifact id of "app-hdfs-binary-compactor"

Step 2a. Configure the Service

In the "Service" tab, configure the parameters:

- **HDFS Base**
    - The base directory on the hdfs file system under which all content will be written
    - Form
        - hdfs://<server>:<port>/<path>
    - Example
        - hdfs://localhost:9000/myBase
- **Content source directory**
    - Add one or more content source directories that will be scanned for compaction. These directories will be located below the base given above
    - Form
        - <directory name>
    - Example
        - aspire_filesystem_source
- **Period**
    - The period between scans of each content source directory
    - Form
        - <number> <unit>
    - Example
        - 12h
- **Threshold**
    - The number of files that must exist in the directory before compaction takes place
- **Suppress deletes**
    - Checking this box will prevent binary files that are in an existing HAR file from being deleted when a delete action is encountered.
- **HDFS Options**
    - **Security**
        - Choose the type of security to use to access the HDFS file system- Kerberos or none
    - **User principle**
        - The principal user for Kerberos
    - **Key tab file**
        - The user's key tab file
        - Form
            - File path
        - Example

- config/myUser.keytab
    - **Add resources**
        - Check this box if you need to add Hadoop resources to the configuration (such as site-core.xml)
        - Resource file
            - The path to a resource file to add to the configuration
            - Form
                - /path/to/file
            - Example
                - config/core-site.xml
    - **Block size**
        - The size of block to be used when accessing the HDFS file system
        - Form
            - <number> <unit>
        - Example
            - 32mb
    - **Buffer size**
        - The size of buffer to be used when accessing the HDFS file system
        - Form
            - <number> <unit>
        - Example
            - 32kb
    - **Replication**
        - The HDFS replication factor
        - Form
            - <number>
        - Example
            - 3
- **Configure lock**
    - Check this box if you want to configure the lock files
    - **Tries**
        - The number of attempts to get a lock file before moving on
        - Form
            - Number
        - Example
            - 3
    - **Retry sleep**
        - The period to wait before attempting to get the lock when a previous attempt has failed
        - Form
            - <number> <unit>
        - Example
            - 15s
    - **Expiry**
        - The period after which a lock is deemed to have expired and will be released when another process attempts to get it.
        - Form
            - <number> <unit>
        - Example
            - 12h
- **Debug**
    - Check this box to enable debug messages

Save the service and it will start automatically