Basic Tutorial (Aspire 2)

Follow this link if you are using Aspire for Elasticsearch

The following tutorial gets you started with Aspire in 20 minutes or less. This tutorial is for Aspire 2, if you need a tutorial for an older version (1.x), use the link Getting Started Aspire 1.x This tutorial will take you through installing Aspire, an Aspire connector application, and an Aspire publisher application (that simply writes to a file instead of to a search engine for indexing). It will give you an idea of how you install Aspire applications using the System Admin user interface (no programming knowledge needed).

There is another quick start tutorial to learn how to build Aspire distributions from scratch using Maven prototypes and Maven component repositories. Quick Start with Distribution Archetype

Additionally, there are also tutorials for each type of connector application you might wish to install; those are located under the section for each Connector.

Prerequisites

Before you begin, you need to be registered to use Aspire (go to http://aspire.searchtechnologies.com/) if you haven't already done that.

You will need your user registration name and password in order to complete this tutorial.

Step 1: Install Java

The version of Java you should use depends on the Aspire version you are targeting to:

- Aspire 2.1.2 and earlier runs on Java 1.6 or Java 1.7
- Aspire 2.2 and up requires to run at Java 1.7.

Note that we recommend installing the Java JDK (Java Development Kit), just in case you want to create your own Aspire Components in the future. But really, only the JRE (Java Runtime Environment) is absolutely required.

1. Download and install the latest version of the Java JDK appropriate for the system that will run Aspire: http://java.com/en/download/manual.jsp
    - If you have a 64 bit machine, we recommend installing the 64 bit version of Java. That will allow you to create large-memory instances of Aspire.
        - The Aspire framework itself does not use up that much memory (100mb or so). But some applications may store big hash tables to improve performance, so it's best to have the 64 bit JVM (Java Virtual Machine), just in case you need it someday.
2. Test that you can access the "java" command from your console.
    a. Open up a new DOS command-shell (go to the Start menu, enter "cmd" in the "Run" or "Search for Programs" field, and then execute the cmd.exe program).
    b. At the prompt, enter the following, then press the Enter key: java -version
    c. Success is indicated when version information is returned.

up to Aspire 2.1.2:

```
> java -version
java version "1.6.0_18"
Java(TM) SE Runtime Environment (build 1.6.0_18-b07)
Java HotSpot(TM) 64-Bit Server VM (build 16.0-b13, mixed mode)
```

or as of Aspire (2.2 Release) :

```
> java -version
java version "1.7.0_79"
Java(TM) SE Runtime Environment (build 1.7.0_79-b15)
Java HotSpot(TM) 64-Bit Server VM (build 24.79-b02, mixed mode)
```

Step 2: Download the Quick-Start Distribution

Download and unpack https://wiki.searchtechnologies.com/binaries/. Select the most recent Aspire binary (aspire-distribution-X.X.zip). For purposes of this tutorial, we'll use "aspire-quick-start" as the directory name to which you unpack Aspire.

Note: This is the best way to evaluate Aspire with or without premium connectors. The preferred method if you are actually building a distribution is to use the Distribution Archetype, which requires also downloading a Maven client.

The download will create a directory structure similar to that described in Aspire Directory Structure.

Step 3: Edit the Aspire settings.xml File

Go to the directory where you unpacked Aspire (such as "aspire-quick-start") and go to the configuration directory /config. Open the settings.xml file with a text or XML editor. Look for the maven repository tag. You need to replace the user name and password that displays with the user name and password you used to register for Aspire.

```
<repository type="maven">
     <defaultVersion>2.2</defaultVersion>
     <remoteRepositories>
       <remoteRepository>
         <id>stPublic</id>
         <url>
           http://repository.searchtechnologies.com/artifactory/simple/community-public/
         </url>
         <user>YOUR-REGISTERED-USERNAME</user>
         <password>YOUR-REGISTERED-PASSWORD</password>
       </remoteRepository>
     </remoteRepositories>
   </repository>
```

Once you've entered your user name and password, save and close the file. Make sure you save it so it remains a .xml file

Step 4: Start Up Aspire

First, make sure you have access to the internet so that Aspire can download components. Next, still in the Aspire directory you created, change to the /bin directory and type "startup" to launch Aspire.

Note that "startup" is a batch script (on Windows) or a shell script (on Unix) that can be modified as necessary if you need more memory or need to set other system properties.

Aspire may take a few seconds to load all of the necessary components. You will see feedback to the command prompt during the startup.

NOTE: If you are downloading Aspire Community, ignore the error message about being unable to download the com.searchtechnologies.aspire: aspire-dcm-enterprise component. The aspire-dcm-enterprise component is available only with Enterprise systems (and is used for Distributed Processing).

Step 5: Launch Aspire and open the Content Source Management Page

blocked URL

Launch Aspire (if it's not already running). See:

- Launching Aspire

Browse to: http://localhost:50505. For details on using the Aspire Content Source Management page, please refer to UI Introduction.

Step 6: Add a new File System Content Source

blocked URL

To specify exactly what File System folder to crawl, we will need to create a new "Content Source".

To create a new content source:

1. From the Aspire 2 Home page, click on "Add Source" button.
2. Click on "File System Connector".

Step 6a: Specify Basic Information

blocked URL

General Configuration Tab

In the "General" tab in the Add New Content Source window, specify basic information for the content source:

1. Enter a content source name in the "Name" field.

   This is any useful name which you decide is a good name for the source. It will be displayed in the content source page, in error messages, etc.

2. Click on the "Active?" checkbox to add a checkmark.

   Unchecking the "Active?" option allows you to configure content sources but not have them enabled. This is useful if the folder will be under maintenance and no crawls are wanted during that period of time.

3. Click on the "Schedule" drop-down list and select one of the following: *Manually, Periodically*, *Daily,* or *Weekly.*

Aspire can automatically schedule content sources to perform incremental crawls on a set schedule, such as once a day, several times a week, or periodically (every N minutes or hours).
For the purposes of this tutorial, you may want to select *Manually* and then set up a regular crawling schedule later.

4. After selecting a *Schedule* type, specify the details, if applicable:
    a. *Manually*: No additional options.
    b. *Periodically*: Specify the "Run every:" options by entering the number of "hours" and "minutes."
    c. *Daily*: Specify the "Start time:" by clicking on the hours and minutes drop-down lists and selecting options.
    d. *Weekly*: Specify the "Start time:" by clicking on the hours and minutes drop-down lists and selecting options, then clicking on the day checkboxes to specify days of the week to run the crawl.
    e. *Advance*: Enter a custom CRON Expression (e.g. 0 0 0 ? * *)

Step 6b: Specify the Connector Information

blocked URL

In the "Connector" tab, specify the connection information to crawl the File System folder.

1. Enter the folder path you want to crawl.

    For **Windows**: Use the following format D:\folder\folder1\
    For **Linux**: Use the following format /home/user/folder/folder1/

2. Check on the other options as needed:
    a. Index folders?: index subfolders as items. If unchecked, only files will be indexed.
    b. Scan subfolders?: Scan through subfolder's child nodes.
    c. Include/Exclude patterns: Enter regex patterns to include or exclude files/folders based on URL matches.

Step 6c: Specify Workflow Information

blocked URL

In the "Workflow" tab, specify the workflow steps for the jobs that come out of the crawl. Drag and drop rules to determine which steps each AspireObject should follow after being crawled. At a minimum there will be a rule for where to publish the document. This is also where the rules for any transformations on the data before sending it to a search engine are established. See Workflow for more information.

1. For the purpose of this tutorial, drag and drop the *Publish To File* rule found under the *Publishers* tab to the **onPublish** Workflow tree.
    a. Specify a *Name* and *Description* for the Publisher.
    b. Leave the default values for the number of logged jobs and the file name.
    c. Click *Add*.

After completing this steps click on the Save button and you'll be sent back to the Home Page.

Step 7: Initiate the Full Crawl

blocked URL

Now that the content source is set up, the crawl can be initiated.

1. Click on the Full crawl option (lower left of content source card) to initiate the crawl.

During the Crawl

blocked URL

During the crawl, you can do the following:

- Click on the "Refresh" button on the Content Sources page to view the latest status of the crawl.

  The status will show **RUNNING** while the crawl is going, and **CRAWLED** when it is finished.

- Click on "Complete" to view the number of documents crawled so far, the number of documents submitted, and the number of documents with errors.

If there are errors, you will get a clickable "Error" flag that will take you to a detailed error message page.

Step 8: Congratulate yourself! (and shutdown)

Congratulations!!

You have completed the 20-minute quick start. We hope it was a fun experience.

To shutdown Aspire, go to the home page (http://localhost:50505/aspire) and click on the "Shutdown" button (that's the red button to the right of the server name). Or, you could go to the Aspire console window (where you started Aspire with "bin\startup") and type "shutdown" and then press the Return or Enter key. Either way will shut down Aspire.

Cheers!

The Aspire Development Team