

# Database Server How to Configure

## On this page

- [Step 1. Launch Aspire and open the Content Source Management page](#)
- [Step 2. Add a new Database Server content source](#)
  - [Step 2a. Specify basic information](#)
  - [Step 2b. Specify the connector information](#)
  - [Step 2c. Specify workflow information](#)
  - [During the Crawl](#)
  - [Exclude and Include File](#)

## Step 1. Launch Aspire and open the Content Source Management page

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

- See [Launch Control](#)
- Browse to: <http://localhost:50505>

For details on using the Aspire Content Source Management page, refer to [Admin UI](#)

## Step 2. Add a new Database Server content source

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

To create a new content source:

1. From the Content Source , click on "Add Source" button.
2. Click on "Database Server Connector".

### Step 2a. Specify basic information


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


1. Enter a content source name in the "Name" field.
  - a. 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 **Scheduled** pulldown list and select one of the following: *Manually*, *Periodically*, *Daily*, *Weekly* or *Advanced*.
  - a. Aspire can automatically schedule content sources to be crawled 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.
3. Click on the **Action** pulldown list to select one of the following: *Start*, *Stop*, *Pause*, or *Resume*.
  - a. This is the action that will be performed for that specific schedule.
4. Click on the **Crawl** pulldown list and select one of the following: *Full*, *Real Time*, or *Cache Groups*.
  - a. This will be the type of crawl to execute for that specific schedule.


After selecting a Scheduled, specify the details, if applicable:

- *Manually*: No additional options.
- *Periodically*: Specify the "Run every:" options by entering the number of "hours" and "minutes."
- *Daily*: Specify the "Start time:" by clicking on the hours and minutes drop-down lists and selecting options.

- **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.
- **Advanced:** Enter a custom CRON Expression (e.g. 0 0 0 ? \* \*)

 You can add more schedules by clicking in the **Add New** option, and rearrange the order of the schedules.

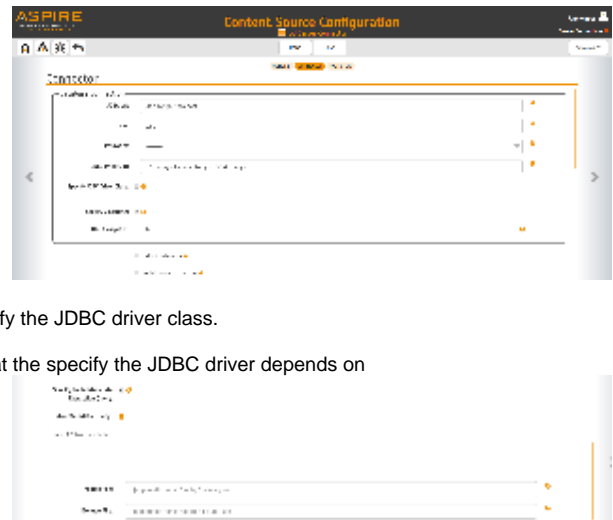
 If you want to disable the content source just unselect the the "Enable" checkbox. This is useful if the folder will be under maintenance and no crawls are wanted during that period of time.

 Real Time and Cache Groups crawl will be available depending of the connector.

## Step 2b. Specify the connector information

In the "Connector" tab, specify the connection information to crawl the Database Server.

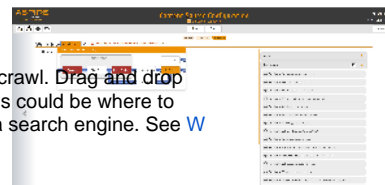
1. **JDBC Url:** The default JDBC url for your RDBMS server and database. For example, "jdbc:mysql://server:3306/mydb" (MySQL). This will vary depending on the type of RDBMS.
2. **User:** The default name of a database user with the access to all of the databases and tables containing data to be indexed.
3. **Password:** The password for the database user.
4. **JDBC Driver Jar:** Path to the default JDBC driver jar file for your RDBMS. Typically this is placed in the "lib" directory inside your Aspire Home, for example "lib/myjdbcdriver.jar".
5. **Specify JDBC Driver Class:** Check this box if you need to manually specify the JDBC driver class.
  - a. **JDBC Driver Class:**
6. **Specify Classpath:** Check this box if you need to add external jar files that the specify the JDBC driver depends on
  - a. **JDBC Classpath:** The class path for external jar files.
7. **Index Containers:** Select if databases and tables are to be indexed.  
Clear to index rows only.
8. **Enable rows extraction:** Select if the tables should be scanned.
9. **Extract table row count:** Select if the row count of the tables should be extracted.
10. **Specify Table Metadata Extraction Query:** Check this to specify a query to extract the tables object metadata
  - a. **Table Metadata Query:** The query to run for extracting the tables metadata. The placeholders {{database}}, {{schema}} and {{table}} can be used.
11. **Include File:** Use this field to specify the path of the include file. The file will be used to filter databases and tables from the crawl.
12. **Exclude File:** Use this field to specify the path of the exclude file. The file will be used to exclude databases and tables from the crawl.



## Step 2c. Specify workflow information

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 should an item follow after being crawled. This rules could be where to publish the document or transformations needed on the data before sending it to a search engine. 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. Click **Add**.



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

## During the Crawl

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.

## Exclude and Include File

The exclude and include can be used to filter the crawled databases and tables. It has the following structure:

```
{
  "database" : [
    {
      "name" : "dbtest1",
      "pattern" : false
    }, {
      "name" : ".*2",
      "pattern" : true
    }
  ],
  "table" : [
    {
      "name" : "test1",
      "database" : "dbtest1",
      "pattern" : false
    }, {
      "name" : ".*3",
      "pattern" : true
    }, {
      "name" : ".*4",
      "database" : "dbtest2",
      "pattern" : true
    }
  ]
}
```

1. **database:** List of databases to exclude/include from the crawl.
  - a. **name:** name or pattern of the database to exclude/include.
  - b. **pattern:** specifies if the name parameter is a regex pattern or not.
2. **table:** List of tables to exclude/include from the crawl.
  - a. **name:** name or pattern of the table to exclude/include.
  - b. **database:** optional - name of the database that contains the table to exclude/include. If specified only the tables contained by the database will be filtered
  - c. **pattern:** specifies if the name parameter is a regex pattern or not.