

Aspider FAQ & Troubleshooting

FAQs

General

Why does an incremental crawl last as long as a full crawl?

Some connectors perform incremental crawls based on snapshot entries, which are meant to match the exact documents that have been indexed by the connector to the search engine. On an incremental crawl, the connector fully crawls the repository the same way as a full crawl, but it only indexes the modified, new or deleted documents during that crawl.

For a discussion on crawling, see [Full & Incremental Crawls](#).

Save your content source before creating or editing another one

Failing to save a content source before creating or editing another content source can result in an error.

```
ERROR [aspire]: Exception received attempting to get execute component command com.searchtechnologies.aspire.services.AspireException: Unable to find content source
```

Save the initial content source before creating or working on another.

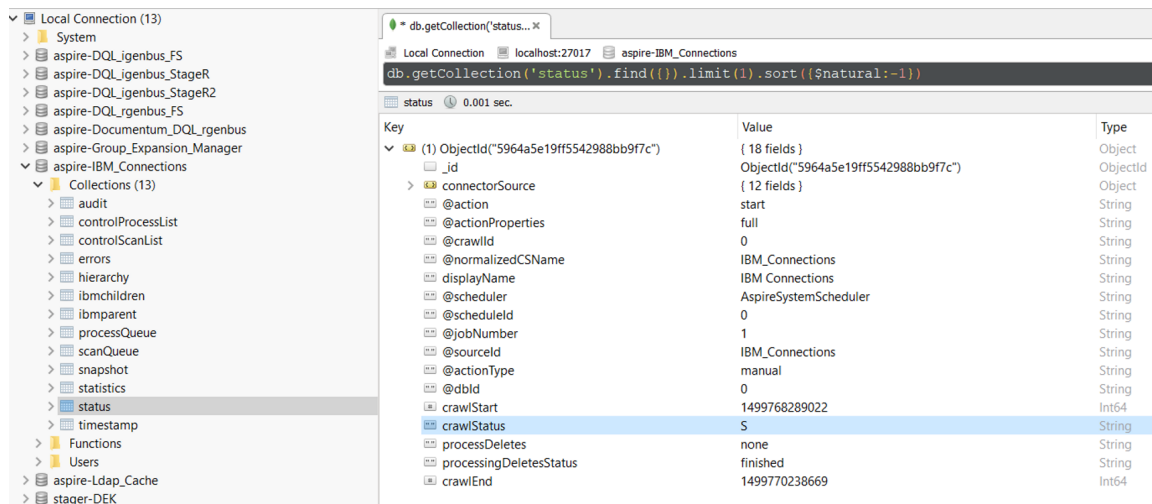
My connector keeps the same status "Running" and is not doing anything

After a crawl has finished, the connector status may not be updated correctly.

To confirm this, do the following:

1. In RoboMongo, go to your connector database (like: *aspire-nameOfYourConnector*).
2. Open the "Status" collection and perform the following query:

```
db.getCollection('status').find({}).limit(1).sort({$natural:-1})
```



The screenshot shows the RoboMongo interface. On the left, the database structure is visible, with 'Collections (13)' expanded under 'aspire-IBM_Connections'. The 'status' collection is selected. On the right, the query `db.getCollection('status').find({}).limit(1).sort({$natural:-1})` is entered and executed. The result is a single document with the following fields:

Key	Value	Type
<code>_id</code>	<code>ObjectId("5964a5e19ff5542988bb9f7c")</code>	ObjectId
<code>connectorSource</code>	<code>{ 18 fields }</code>	Object
<code>@action</code>	<code>start</code>	String
<code>@actionProperties</code>	<code>full</code>	String
<code>@crawlId</code>	<code>0</code>	String
<code>@normalizedCSName</code>	<code>IBM_Connections</code>	String
<code>displayName</code>	<code>IBM Connections</code>	String
<code>@scheduler</code>	<code>AspireSystemScheduler</code>	String
<code>@scheduleId</code>	<code>0</code>	String
<code>@jobNumber</code>	<code>1</code>	String
<code>@sourceId</code>	<code>IBM_Connections</code>	String
<code>@actionType</code>	<code>manual</code>	String
<code>@dbId</code>	<code>0</code>	String
<code>crawlStart</code>	<code>1499768289022</code>	Int64
<code>crawlStatus</code>	<code>S</code>	String
<code>processDeletes</code>	<code>none</code>	String
<code>processingDeletesStatus</code>	<code>finished</code>	String
<code>crawlEnd</code>	<code>1499770238669</code>	Int64

3. Edit the entry and set the status to "S" (Completed).

```

Edit Document
localhost:27017 aspire-IBM_Connections status


{
  "_id" : ObjectId("5964a5e19ff5542988bb9f7c"),
  "connectorSource" : {
    "IBMServer" : "https://ws8-ibm5.qa.local/",
    "IBMUser" : "wasadmin",
    "IBMPassword" : "encrypted:90927FC87FB8745A7A68BD76DC96A96B",
    "useLTPA" : "false",
    "IBMLoginUrl" : null,
    "pageSize" : "100",
    "extractACL" : "true",
    "ldapComponent" : "/Ldap_Cache",
    "ldapServerUrl" : "myldapValue",
    "ldapUserName" : "sams",
    "crawlByApps" : "false",
    "withLimitedAccess" : "false"
  },
  "$action" : "start",
  "$actionProperties" : "full",
  "$crawlId" : "0",
  "$normalizedConnName" : "IBM_Connections",
  "displayName" : "IBM_Connections",
  "$scheduler" : "AspireSystemScheduler",
  "$scheduleId" : "0",
  "$jobNumber" : "1",
  "$sourceId" : "IBM_Connections",
  "$actionType" : "manual",
  "$dbId" : "0",
  "crawlStart" : NumberLong(1499768289022),
  "crawlStatus" : "S",
  "processDeletes" : "none",
  "processingDeletesStatus" : "finished",
  "crawlEnd" : NumberLong(1499770238669)
}

```

Note: To see the full options of "Status" values, see [MongoDB Collection Status](#).

My connector is not providing group expansion results

Make sure your connector has a manual scheduler configured for Group Expansion.

 Add New

Scheduled:

Manually

Action:

Start

Crawl:

Cache Groups

1. Go to the Aspire [debug console](#), and look for the respective scheduler (in the fourth table: Aspire Application Scheduler).

Aspire Application Scheduler:

Scheduler	enabled				
Name	Schedule	Last run	Next run	Status	
IBM_Connections:1	manual	never	disabled	disabled	detail run
IBM_Connections:2	manual	never	disabled	disabled	detail run

2. If you are unsure which scheduler is for Group Expansion, you can check the Schedule Detail.

- You can identify it with the value: cacheGroups

Schedule Detail
[X]

ID	1
Schedule ID	aspire.AspireSystemScheduler.1
Name	IBM_Connections:1
Schedule	manual
Source ID	IBM_Connections
Event	start
Properties	cacheGroups
Enabled	false
Singleton	true


Job

```

<doc action="start" actionProperties="cacheGroups" crawlId="0" normalizedCSNames="IBM_Connections">
  <connectorSource>
    <IBMServer>https://vs8-ibm5.qa.local/</IBMServer>
    <IBMUser>vsadmin</IBMUser>
    <IBMPassword>encrypted:9D927FC87FB8745A7A688D76DC96A96B</IBMPassword>
    <useLTPA>false</useLTPA>
    <IBMLoginUrl/>
    <pageSize>100</pageSize>
    <extractACL>true</extractACL>
    <ldapComponent>/Ldap_Cache</ldapComponent>
    <ldapUserGUID>myGUIDValue</ldapUserGUID>
    <ldapUseNameSpace>true</ldapUseNameSpace>
  </connectorSource>
</doc>

```

3.To run the Group Expansion process, click **Run**.

 **Aspire Application Scheduler:**

Scheduler	enabled				
Name	Schedule	Last run	Next run	Status	
IBM_Connections:1	manual	never	disabled	disabled	detail run
IBM_Connections:2	manual	never	disabled	disabled	detail run

Authentication Issues

(Cookie Based) I have added the "username" and "password" fields but I can't still authenticate.

Sometimes just the username and password fields are not enough to authenticate to a site. Some sites requires some custom fields or even the "submit" button in order to successfully authenticate you. So you may have to add them as custom fields in the Aspider Configuration.

You can also open the browser inspect mode in order to break down the authentication request and make sure you are not missing any field.

(Cookie Based) How to include those dynamic "hidden" elements into the Aspider authentication request?

Don't worry about those hidden fields inside the form, Aspider will automatically include them in the request, you don't have to do anything.

(Cookie Based) My initial login is successful but shortly after, Aspider can't connect to already discovered URLs.

Watch out for "logout" pages which usually send requests to the browsers to clear their cookies. If Aspider is requested to clear its cookies for logging out, it will do that and will not try to re-login.

Suggestion: Add an exclusion pattern for the "log out" pages.

- If "Scan Excluded Items" is selected, make sure the "Do not follow patterns" also contains a pattern that matches the "log out" page.

(Cookie Based) My site requires two different HTTP requests in order to authenticate, can Aspider handle that?

Unfortunately not at the moment, Aspider Cookie Based Authentication is built to send only one request, but we are already considering improvements for it.

(Cookie Based) Can Aspider handle javascript based authentications?

If your login form relies on javascript for sending the Authentication request your form element probably won't have an "action" attribute, which Aspider use for sending the POST request to. So you wouldn't be able to authenticate.

(Cookie Based) Which versions of SSL are supported by the Aspider web crawler?

As of Java 8, Aspider supports and has been tested to work on the following protocols:

- TLSv1
- TLSv1.1
- TLSv1.2

Note: *SSLv2 and SSLv3 are not supported by Aspider.*