


Azure Data Lake FAQ & Troubleshooting

FAQs

Specific

 No available questions at this moment

General

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

Some connectors perform incremental crawls based on snapshot files, 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 file system 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.

Troubleshooting

 com.searchtechnologies.aspire.services.AspireException: Unable to access path: '/test/NOACCESS'. Missing READ and EXECUTE access. Please check your application created. Skipped
at com.searchtechnologies.aspire.components.AzureDataLakeStoreQueryManager.getEnumerateDirectory(AzureDataLakeStoreQueryManager.java:79)
at com.searchtechnologies.aspire.components.AzureDataLakeStoreRAP.scan(AzureDataLakeStoreRAP.java:167)
at com.searchtechnologies.aspire.connector.framework.ScanItem.process(ScanItem.java:87)
at com.searchtechnologies.aspire.application.groovy.ScriptingStage.process(ScriptingStage.java:245)
at com.searchtechnologies.aspire.application.groovy.GroovyJob.or(GroovyJob.java:567)
at com.searchtechnologies.aspire.application.groovy.GroovyJob\$or.call(Unknown Source)
at Script1.run(Script1.groovy:18)
at com.searchtechnologies.aspire.application.Pipeline.runScript(Pipeline.java:489)
at com.searchtechnologies.aspire.application.groovy.GroovyJobHandler.runNested(GroovyJobHandler.java:64)
at com.searchtechnologies.aspire.application.JobHandlerImpl.run(JobHandlerImpl.java:91)
at java.util.concurrent.ThreadPoolExecutor.runWorker(Unknown Source)
at java.util.concurrent.ThreadPoolExecutor\$Worker.run(Unknown Source)
at java.lang.Thread.run(Unknown Source)

Problem

Folder not able to be scanned, due to access rights.

Solution

Make sure READ and EXECUTE access (r-x) is set properly to the Application created.



```
com.searchtechnologies.aspire.services.AspireException: AzureDataLakeStoreQueryManager.getDirectoryOrFileEntry: Path not found or  
access error: '/doesnotexist'. Skipped  
at com.searchtechnologies.aspire.components.AzureDataLakeStoreQueryManager.getDirectoryOrFileEntry  
(AzureDataLakeStoreQueryManager.java:116)  
at com.searchtechnologies.aspire.components.AzureDataLakeStoreRAP.processCrawlRoot(AzureDataLakeStoreRAP.java:47)  
at com.searchtechnologies.aspire.connector.framework.ProcessCrawlRoot.process(ProcessCrawlRoot.java:69)  
at com.searchtechnologies.aspire.application.groovy.ScriptingStage.process(ScriptingStage.java:245)  
at com.searchtechnologies.aspire.application.groovy.GroovyJob.or(GroovyJob.java:567)  
at com.searchtechnologies.aspire.application.groovy.GroovyJob$or.call(Unknown Source)  
at Script3.run(Script3.groovy:35)  
at com.searchtechnologies.aspire.application.Pipeline.runScript(Pipeline.java:489)  
at com.searchtechnologies.aspire.application.groovy.GroovyJobHandler.runNested(GroovyJobHandler.java:64)  
at com.searchtechnologies.aspire.application.JobHandlerImpl.run(JobHandlerImpl.java:91)  
at java.util.concurrent.ThreadPoolExecutor.runWorker(Unknown Source)  
at java.util.concurrent.ThreadPoolExecutor$Worker.run(Unknown Source)  
at java.lang.Thread.run(Unknown Source)
```

Problem

Folder configured on scanning settings does not exist.

Solution

Revisit connection sources on Azure Data Lake connector and double check the list of sources.