Aspire 4.0 Release Notes

> ⓘ **For version 4.0, Aspire requires a license file to run.**
>
> **See Aspire Licensing for information on obtaining a license.**

The following are the NoSQL DB providers supported by the Aspire 4.0 release:

- Elasticsearch version 7.1.1
- MongoDB version 4.0

The supported version of StageR is 1.2 and it works with MongoDB v. 3.4.10.

Below you can find the list of updates included in this version.

New and Enhanced Features

Aspire Core and Framework Components

- New Elasticsearch as a NoSQL provider.
- All Publishers are using the new Publisher Framework.
- Updates to the Listener service (push updates).
- Cluster Mode improvements (Zookeeper stability).
- Implemented the FIFOQueue for the MongoDB Provider.
- New options for the Extract Text configuration and new Throttling section in Advanced Configuration for connectors.
- Background processing.
- Saga natural language processing is available as an Aspire plug-in and can be used to perform NLP as part of Aspire workflow.

Aspire UI

- Import/Export System configuration.
- Log Browser.

Connectors

- OneDrive.

Publishers

- Amazon S3.
- Background Queue.
- Azure Search.

Plugins

- Stager BDC Plugin.
  - Now supports Sharepoint 2019.

Services

- Binary Store.
- Thumbnails.

Bug Fixes

Aspire Core and Framework Components

- Missing headers on OAuth classes.
- Wrong URL info for the Aspire UI Authentication documentation in settings.xml file.
- Master password ssh file not working on Centos OS.
- Errors processing failed documents with the Exception Patterns option.
- Components on Workflow not saved if the content source was not saved first.
- Invalid entitlements host caused missing workflow applications.
- Double click ignored on disabled workflow item.
- NPE after shutting down 2 Aspire instances in distributed mode.
- Java 1.8 Error when name of the Application and name of Publish was the same.
- Aspire not starting in shell mode on Centos OS.
- Publisher added to the workflow not being unpacked into cache folder so they were unavailable and not working.
- Error installing Aspire as a service in Windows.
- Status not displayed in Aspire after a crawl was aborted.
- NPE scheduling the "cacheGroups" option without the GEM configured.
- Two different entries in status collection being generated for the same crawl ID.
- NPE having Artifactory user with not entitlements assigned.
- Mongo database name limit exceeded by the Aspire Database name.
- NPE with the Non Text Document Filter and Open Data Stream options enabled.
- NPE using encrypted password at the SSL settings in settings.xml file.
- NPE pausing a crawl with MongoDB and Zookeeper in distributed mode.
- Error trying to import a Service since some services do not have a workflow associated.
- Previous crawl errors displayed when current crawl was running.
- Crawls on distributed mode not populating correctly ancestor ID and ACLs.
- Error uninstalling Aspire as a service.
- Aspire not getting alert if Elasticsearch provider is not running.
- Crawl statistics not reflecting the deletes if there were adds/updates.
- NPE after an authentication method configured in the settings.xml file.
- NPE displayed while stopping a crawl after it just started.
- Some Aspire UI settings configured in settings.xml file being ignored.
- Out of Memory error using a very big number in Hierarchy Cache Size option.
- Every time a groovy script was updated, a blank line was added at the beginning of the script.
- Invalid characters validation in the Extension List option of the Non Text Document filter.

Applications

- Archive Extractor
  - Using Select/Deselect All option closed the Configuration window.
- AVRO Extractor
  - ASPIRE-8112/ASPIRE-8113 Routing section options not displaying correctly.
- Hierarchy Extractor
  - The User/Group field on ACLs section is now required.

Aspire UI

- Typos on Accenture license information.
- UI refreshing stacks over and over while changing between the Cards View and the List View.
- Aspire DXF not accepting Windows relative paths.
- The word "content sources" displayed in the Service Group control.
- Navigation controls at the bottom overlapping the footer.
- Fixed special characters allowed in the connector's name.

Connectors

- Adobe Experience Manager
  - Use scheduled (de)activation item settings not working without include/exclude properties.
  - Fetch ACLs option not working.
  - Updates on pages not crawled on incremental.
  - Wrong credentials threw unclear message on Basic Authentication.
  - Malformed URLs not validated.
  - More user friendly exception for non-existent pages/assets.
  - Normalized date format for the "lastModified" field.

- Amazon S3
  - Crawl failing for items published with the S3 Publisher.
  - Some exceptions using the connector, the Archive Extractor application and the Elasticsearch publisher.
  - Crawl failing if directory URL not ending in the "/" character.
  - Using bad Include Pattern prevented crawl to start.

- Aspider
  - Crawls not finishing in distributed mode.
  - Updates processed as Add instead of Updates.
  - Missed some options on the Extract Text section.
  - Hierarchy info appearing having the Hierarchy option disabled.
  - NPE displayed while content cleanup is selected but nothing is configured.
  - A [NO-NAME] value displayed in the hierarchy section.
  - Extract Text & Hierarchy options took out of the Advanced Configuration section.
  - Content cleanup of web pages not working in Aspire 3.3.0.4.

- Images not being crawled using the Extract Text option enabled.

- Azure Blob
  - Seed file option is not working.
  - <Non text document> tag not published using open data stream.
  - Split Words per XML/HTML Tag is not working.
  - HTML Output not producing any document output.
  - Crawl errors displayed in the UI.
  - Incremental crawl not detecting updates.
  - Storage Connection String set as a placeholder.
  - Problem crawling folders.

- Azure EventHub
  - Valid tooltips for the field in the Credentials section.

- Box
  - SSLExceptions during crawls (HTTP error code 429)
  - Incremental actions not working.
  - Acces token issue during crawls.
  - Incremental crawls getting more items than expected.

- Database Server
  - Scan errors crawling all tables in the RDBMS.

- Elasticsearch
  - 429 Error Management.

- File System
  - Hierarchy information incomplete.
  - No error using invalid filename specified in the "Path to Root directories file" field.
  - NPE using Multiple starting points option.

- IBM Connections
  - Crawling specific endpoints (Communities, Forums, and Wikis) not working.
  - Error caching groups.
  - Incremental after deletes using Elasticsearch not working.
  - Querying an IBM Domino getting an OperationNotSupportedException (LDAP: error code 12 - Unavailable Critical Extension).
  - Hierarchy information not being published to Elasticsearch.
- Lotus
  - Error publishing to Elasticsearch. Check **Known Issues** section for workaround for this issue.

- RDB Snapshot
  - Problems crawling delete actions.

- RDB Tables
  - Exception using the Slices option not reported on the UI.

- Sharepoint 2013
  - Connector processes same document with different ID between crawls.
  - NPE pausing a crawl.
  - Problem running incremental using Lists option. All content being crawled.

- Sharepoint 2016
  - Issue on incremental when an External List was included in a pattern.
  - Problem on incremental using Tokens with Crawl Attachments option enabled.
  - Connector not crawling folders created under site collection.

- Sharepoint Online
  - List threshold: not all items on a big list are being crawled.
  - Group Expansion not working.
  - Error generating FetchUrl and Display URL for link list items inside a folder.

- SMB
  - Crashes with start URL ending without slash
  - Connector failing when crawling specific file url and regex
  - Deny permissions are missing
  - SMB doesnt' detect ACL changes on incremental crawl
  - SMB connector override last access date only when it was changed by the connector

- ServiceNow
  - Use Agregate API not working.
  - ID Unexpected displayed instead of ACL.
  - Group expansion not being checked by default.

- StageR
  - No error message on console or UI indicating wrong storage/scope used.

- Yammer
  - No error message on Aspire Web UI when Yammer token is invalid.

Publishers

- Elasticsearch
    - Updated items published as a new item.
    - NPE when using an incorrect ES port/host.
    - Added validation for malformed index name.
    - Groovy Transform is not validated with absolute/relative path.
    - Minor UI changes (tooltips and validations)

- Google Cloud Search
    - NPE when hierarchy info not coming from the connector.
    - $superSearcherAcl being added as part of the ACLs when setting is empty.
    - Content type Raw not extracting the content for binary files.
    - Option to populate the gcsUniqueId field.
    - SocketTimeout exception.
    - Date fields using a month range from 0 to 11 instead of 1 to 12.

- Solr
    - Option to set multiple URLs not working.
    - XSL Transform is not validated with absolute/relative path.
    - Solr URL field required Malformed URL validation.
    - Removed info from tooltip about 'default core'. Core field now is required.

- StageR
    - Delete All Action is not always executed first.

Services

- NPE while using services with workflows.
- Group Expansion not loading after Aspire was restarted.
- Error while adding Services with no workflows.
- Broken images/icons on Services UI.
- Encryption issue with authentication using LDAP Cache Service.
- LDAP Cache Service: Unavailable Critical Extension error querying IBM Domino.
- LDAP Cache error after importing the Service and run it.
- LDAP Cache authentication problem using service account.
- Discovery by Regex will throw error for non-pst files.

Known Issues

Aspire Core and Framework Components

- Completed items not being removed from the process queue.
- Crawl time execution still running after pause it.
- Felix startup warning using Java version 11.
- Connectors/publishers saved twice when Aspire components are still downloading.
- HttpFeeder - Servlet added with the same name of another servlet is not notified in the UI.
- Error validating field Maximum size on Extract Text

Publishers

- Elasticsearch: Error publishing hierarchy information to the index using the Lotus connector. This will be fixed for Aspire 4.0.1. Meanwhile and as workaround change the transform.groovy file in the line 225 to the following line:
    - ancestors?.getChildren().each() { ancestor ->

External Technical Limitations

- Aspire Core and Framework Components
    - Elasticsearch Provider - "FATAL: Flushing-Error" can happen in some connectors.
- Publisher

    - S3: Current implementation has a limit of 5GB when upload.