Aspire 3.3 Release Notes

For version 3.3, 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 3.3 release:

- MongoDB version 3.6
- HBase version 1.2.4

The supported version of Elasticsearch is 6.3.0

The supported version of StageR is 1.2 ***Note:*** *The latest version of Stager is v. 1.2 and it supports MongoDB v. 3.4.10*

Below you can find a list of the updates for this version.

On this page

New and Enhanced Features

Aspire Core and Framework Components

- Salted Challenge Response Authentication Mechanism (SCRAM) support has been added to the MongoDB used with Aspire.
- The ability to dynamically load jar files has been added to Aspire with Java 9.
- When starting Aspire either normally or in debug mode, the debug line in the `settings.xml` file is handled appropriately.
- A section has been added to the `settings.xml` file for HBase information.
- Logging of remote IP addresses for successful or failed logins will now occur.
- The Mongo provider now encrypts/hashes IDs.
- Record fields have been improved.
- Entitlements checking no longer checks missing components at every restart.
- Time zones have been normalized for Aspire, including logs and statistics.
- The documentation has been updated for Keytab/Kerberos.
- Improvements have been added to Job usage.
- Updates have been made to the *ExtractText* default configuration limit for text extracted from a stream.
- List page retrieval and metadata extraction have been improved in SharePoint Commons.

Aspire UI

- To re-fetch entitled components (after deleting the Resources folder), an "Allow Refresh" button has been added.
- The ability to show Provider Information has been added.

Connectors

- Aspider
  - A Headless browser has been added for rendering dynamically generated pages (client-side JavaScript pages).
- IBM Connections
- Elastic
- SharePoint 2010
  - On the Multiple URIs drop-down, when the 'Site Discovery' option is set, the 'Set List View' option is removed.
- SharePoint Online
  - An NPE at crawl end error could occur if bad credentials were used.
  - Incremental crawls no longer detect containers as updated items.
  - Scan recursively was not working as expected.
- SMB
  - Added DFS support and override last access date of documents
- Twitter

Publishers

- Elasticsearch
  - Case sensitive index names can be handled properly now.
- Google Cloud Search
  - A new Google Cloud Search (GCS) publisher receives content from Aspire connectors and uses the Java Client library to index the content into Cloud Search.
- HBase
  - Content can now be deleted.
  - During a full crawl, the publisher now defaults to clean.
  - When not in file configuration mode, the publisher can now be used without security.
- Publish to StageR
  - Field level help has been added for the special scope $record.

Applications

- The Entitlements Admin application has been updated.

Bug Fixes

Aspire Core and Framework Components

- Admin UI
  - The ability to configure a weekly schedule could cause an error when saving
- Aspire Application
  - *ConfigManager* could log a debug message into {aspire.home}/logs/configmanager.log
  - A problem could occur when editing a custom application in the Admin UI
  - Startup problems could occur using the Staging Publisher
- Connector Framework
  - When stopping and restarting Aspire while the GroupDownload process was running, the group download did not start again
- MongoDB Provider
  - The LDAP Cache could report a MongoDB Duplicate key error
  - Aspider could stop with a MongoDB Duplicate key error
- SharePoint Commons
  - An out of memory (OOM) exception could occur during large crawls
  - Added support for incrementals using Aspire Snapshots on SP
- The Aspire Archetype had "http" rather than "https" repository and entitlement URLs
- Failed to connect to Artifactory with custom keystore. Artifactory certificates were added to the distribution. See: https://contentanalytics.digital.accenture.com/pages/viewpage.action?spaceKey=aspire33&title=Crawling+via+HTTPs
- AspireObject was casting an incorrect numeric type when created from JSON
- The AspireObject *isEmpty* method returned true even if the object had children
- The *processDeletes* (String) was missing a Status page
- The Aspire Connector Framework was not using *shouldScan* during incremental crawls
- When running a full crawl, a "Provider 'encrypted' not installed" message could occur
- The Mongo provider generated an invalid JSON object during document conversion
- Audit logs were incomplete
- For AIP integration, the logout action was not being logged
- Publisher framework *retryDelay*, *retryDelayMultiplier* and *maxRetryDelay* properties were not supported by Dynamic XML Forms (DFX)
- The Aspire-Services jar file was missing a noSQL package
- The "Loading Application" message could display whether a connector was loading or not
- Extract Text
  - Use the Apache Tika SAX Parser for Microsoft documents
- Scheduler
  - The option to create a Cache Groups scheduler was not being displayed

Aspire UI

- A Connector component might not show the actual state of a crawl
- The link that points to the Confluence wiki has been updated

Connectors

- Aspider
  - An authentication form error could occur indicating "Target host is not specified while crawling"
  - Neither NTLM nor ADFS authentication was occurring when a host was specified in the Credentials
  - On any port, the Port field was not working correctly with any value except "-1"
  - A crawl could cause a warning about duplicate IDs in MongoDB
  - To indicate that the Gateway was not working, the exception message in ADFS needed updating
- Confluence
  - ACLs info appeared inside the hierarchy section
  - A batch error could display while publishing to Elasticsearch 6.3.0
- Documentum
  - Exception was being thrown during Group Expansion

- File System
  - Starting Directories in the File option was not working as expected
- IBM Connections
  - The connector needed to use the Aspire GroupExpansion instead of SharePoint Integrated security with an optimized IBM Connections Group Downloader
  - Memory leaks could occur
  - During an incremental crawl, the deletes of Blogs, Wikis and Files were not working
  - The Content crawled from IBM Connection did not contain a last-modified date. The problem was with the date format
- Kafka
  - A "NO-NAME" field could occur
- SharePoint 2010
  - A problem could occur when identifying the site-collections for a WEB-Application
  - When adding a link on a site collection to crawl, [NO-NAME] should not be part of the name attribute in the hierarchy section
  - No error should occur during the incremental crawl for the Blog site collection
  - No errors should occur when crawling a specified list (views included)
- SharePoint 2013
  - When crawling incrementals for an External list, the connector was not picking up the changes
  - When crawling SP2013, errors such as "HTTP Error 400. The size of the request headers is too Long" might occur
  - KeyNotFoundException while trying to check attachments for list with lookup references deleted
  - Crawl a list and the name in the hierarchy of the documents will be displayed as NO-NAME even though the items have title field.
  - The placeholder needed to be changed for the 'Seeds file' field
  - The connector was unable to crawl large lists
- SharePoint 2016
  - An error could occur while crawling site
- SharePoint Online
  - NPE crawling on distributed mode. Random NPE in the item complete callback
  - String index out of range while getting a List display url
  - Error while crawling after a crawl was stopped: Item parent wasn't assigned during crawl
- Standalone Mode
  - When a user added a custom connector, feedback needed to be provided by the Aspire UI
- Staging Repository
  - A global variable was not working when configuring the server in the Staging Repository connector
  - When crawling over multiple documents and publishing at two different scopes, the items published could be duplicated
  - The Stager connection could be broken when running a full crawl

Publishers

- Stager BDC Plugin could randomly fail during the crawls after setup
- Elasticsearch
  - *DeleteByQuery* was not being used with Elasticsearch 6.1.1
- GCS Publisher
  - A resource config/application.xml was missing on the jar file
  - A relative path was not working in the *Credentials Key File* field
  - An error could occur when crawling and publishing to GCS
- Kafka
  - An error could be masked when running a non-batched job
- Publish to Avro
  - Validation needed to be added to the Time Rollover Threshold field
- TLS 1.2 support was needed for the SharePoint Security Pre Trimmer

Services

- Azure Group Expander could refuse to start.
- Group Expansion failed if user data exceeded the Mongo Max Document Limit (16MB)
- For Aspire Distributed mode, Services in the master node were not starting automatically after saving changes
- Errors to reflect failed Services were not being generated
- Services that were set up in an Aspire cluster were not synced up correctly
- Azure Active Directory Group Expander
  - Users were not being removed
- Group Expansion Service
  - The *userGroupCache* map was accessed when the Group Expansion Service was running
- LDAP Cache Service
  - The controls did not display and the Schedule was set to Advanced even if Minutes or Hourly were set
  - Problems with LDAP-Cache component could include: reporting a duplicate key error twice, stopping with a duplicate error, taking too long, and refresh refusing to start. The connector could not look up ACL information in the LDAP-Cache component
- You are now able to check the user's cache for the Azure Group Expander via the Debug console

Applications

- Archive Extractor

    - The "Send delete by query first" option could throw an exception
    - Deleting files inside of an archive file was not handled properly for incremental crawls
- AVRO Extractor
    - During an incremental crawl, a "duplicate key error" message could display

Known Issues

Connectors

- FTP
    - FTP connector is only working with Unix systems and not in Windows
- Twitter
    - Full/Incremental crawls for retweets are not working

Publishers

- Google Cloud Search
    - Bundle location error loading the publisher for the first time
    - NullPointerException publishing with Batch and Content Type Raw options
    - ItemUploadRequest exception
    - Pending required field validation for the 'Indexer Type' field