

# Aspire 3.2 Big Data Release What's New

**Aspire 3.2 – Big Data Release** (the first Aspire release as part of Accenture) includes the successful integration of Aspire into the Hadoop ecosystem as a Cloudera parcel with [Accenture AIP](#). However, you can still use Aspire in stand-alone mode or as a parcel inside of Cloudera.

Find more information on the Cloudera parcel configuration at [Aspire Parcel and Service for Cloudera](#).

A logical consequence of the integration into the Hadoop ecosystem is the support for [HBase](#) for crawl metadata and statistics (previously only MongoDB was supported). This facilitates the use of Aspire as part of Big Data solutions.

Review a relevant use case success [here](#). Extended support necessitated refactoring the connector framework and implementing improvements over the previous Aspire version (3.1.1).

All configuration steps needed to configure HBase for crawl metadata can be found at [HBase Settings](#).

Other features of interest are [Licensing](#) and [User Roles](#). User roles improve security control by separating users into "Developers" and "Administrators" with different roles and permissions over the Aspire configuration.

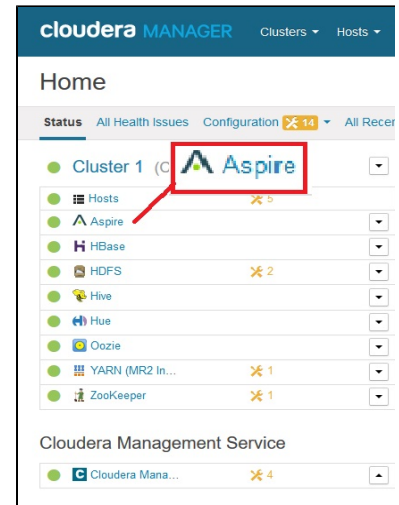
New connectors:

- [HBase Connector](#)
- [Kafka Connector](#)
- [SMB Connector](#)
- [Kinesis](#) IN PROGRESS
- [Exchange](#) IN PROGRESS

New publishers:

- [Publish to HBase](#)
- [Publish to Kafka](#)
- [Solr Cloud](#)

See [Aspire 3.2 Release Notes](#) for more technical information about this release.



## Migrating from Aspire 3.x

When importing a content source from 3.x into 3.2, the following error may occur. The content source may show up with a red "Failed" status.

```

Error message: Unable to start appBundle: com.searchtechnologies.aspire:app-rap-connector
Caused by: com.searchtechnologies.aspire.services.AspireException: Failed to register components from
appBundle: CONTENT_SOURCE_NAME (Parent: <null>)
    at com.searchtechnologies.aspire.application.AspireApplicationImpl.registerAppBundleComponents
(AspireApplicationImpl.java:945)
    at com.searchtechnologies.aspire.application.AspireApplicationImpl.registerAppBundle
(AspireApplicationImpl.java:980)
    at com.searchtechnologies.aspire.application.AspireApplicationComponent.loadApplication
(AspireApplicationComponent.java:696)
    at com.searchtechnologies.aspire.application.AspireApplicationComponent.loadApplication
(AspireApplicationComponent.java:692)
    at com.searchtechnologies.aspire.configuration.ConfigurationManager.reloadApplication
(ConfigurationManager.java:697)
    at com.searchtechnologies.aspire.configuration.ContentSourcesModule.processSyncUnitUpdate
(ContentSourcesModule.java:309)
    at com.searchtechnologies.aspire.configuration.SynchronizedModule.run(SynchronizedModule.java:289)
    at java.lang.Thread.run(Thread.java:748)
Caused by: com.searchtechnologies.aspire.services.AspireException: The value ("${waitForWfApps}") of element
<waitForWfApps> is improperly formatted for a boolean - must be either "true" or "false"
    at com.searchtechnologies.aspire.framework.ComponentImpl.getBooleanFromConfig(ComponentImpl.java:
2634)
    at com.searchtechnologies.aspire.connector.framework.controller.CrawlControllerImpl.initialize
(CrawlControllerImpl.java:260)
    at com.searchtechnologies.aspire.framework.ComponentFactoryImpl.registerComponent
(ComponentFactoryImpl.java:446)
    at com.searchtechnologies.aspire.application.ComponentManagerImpl.registerComponents
(ComponentManagerImpl.java:328)
    at com.searchtechnologies.aspire.application.ComponentManagerImpl.initialize(ComponentManagerImpl.
java:93)
    at com.searchtechnologies.aspire.application.PipelineManagerImpl.initialize(PipelineManagerImpl.java:
75)
    at com.searchtechnologies.aspire.framework.ComponentFactoryImpl.registerComponent
(ComponentFactoryImpl.java:446)
    at com.searchtechnologies.aspire.application.ComponentManagerImpl.registerComponents
(ComponentManagerImpl.java:328)
    at com.searchtechnologies.aspire.application.ComponentManagerImpl.initialize(ComponentManagerImpl.
java:93)
    at com.searchtechnologies.aspire.framework.ComponentFactoryImpl.registerComponent
(ComponentFactoryImpl.java:446)
    at com.searchtechnologies.aspire.application.AspireApplicationImpl.registerAppBundleComponents
(AspireApplicationImpl.java:941)

```

This could happen because Aspire 3.2 connectors contain configuration options that the "content source to import" lack. To fix this error:

1. Click on the content source to access the Configuration page.
2. Click **Save** and **Done**.

Aspire generates the new options and saves them into the configuration files.

## MongoDB Changes

Any migration from Aspire 3.x requires a Full Crawl of all content sources since there was a major refactor on the MongoDB provider component. In specific the following tables changed:

Collection	Fields in 3.x	Fields in 3.2	Compatible
audit	<ul style="list-style-type: none"> <li>• _id (ObjectId)</li> <li>• id (String)</li> <li>• crawlStart (Int64)</li> <li>• url (String)</li> <li>• type (String)</li> <li>• action (String)</li> <li>• batch (String or Null)</li> <li>• ts (Int64)</li> </ul>	<ul style="list-style-type: none"> <li>• _id (ObjectId)</li> <li>• id (String)</li> <li>• ts (Int64)</li> <li>• action (String)</li> </ul>	Yes

errors	<ul style="list-style-type: none"> <li>• _id (ObjectId)</li> <li>• error (Object) <ul style="list-style-type: none"> <li>◦ .@time (Int64)</li> <li>◦ .@crawlTime (Int64)</li> <li>◦ .@cs (String)</li> <li>◦ .@processor (String)</li> <li>◦ .@type (String)</li> <li>◦ ._id (String)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• _id (ObjectId)</li> <li>• message (String)</li> <li>• type (String)</li> <li>• crawlId (String)</li> <li>• time (Int64)</li> </ul>	No
hierarchy	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• itemType (String)</li> <li>• name (String)</li> <li>• ancestors (Object or Null) <ul style="list-style-type: none"> <li>◦ ._id (String)</li> <li>◦ .name (String)</li> <li>◦ .ancestors (Object or Null)</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• itemType (String)</li> <li>• name (String)</li> <li>• ancestors (Object or Null) <ul style="list-style-type: none"> <li>◦ ._id (String)</li> <li>◦ .itemType (String)</li> <li>◦ .name (String)</li> <li>◦ .ancestors (Object or Null)</li> </ul> </li> </ul>	No
processQueue and scanQueue	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• metadata (Object)</li> <li>• type (String)</li> <li>• status (String)</li> <li>• action (String)</li> <li>• timestamp (Int64)</li> <li>• signature (String)</li> <li>• processor (String)</li> <li>• shouldScan (Boolean)</li> <li>• shouldProcess (Boolean)</li> <li>• crawlRetries (Int32) *</li> <li>• name (String)</li> <li>• isCrawlRootItem (Boolean)</li> <li>• hierarchyId (String)</li> <li>• inCrawlRetries (Int32) *</li> </ul>	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• metadata (Object)</li> <li>• url (String)</li> <li>• type (String)</li> <li>• status (String)</li> <li>• action (String)</li> <li>• timestamp (Int64)</li> <li>• signature (String)</li> <li>• processor (String)</li> <li>• shouldScan (Boolean)</li> <li>• shouldProcess (Boolean)</li> <li>• crawlRetries (Int32)</li> <li>• name (String)</li> <li>• isCrawlRootItem (Boolean)</li> <li>• hierarchyId (String)</li> <li>• inCrawlRetries (Int32)</li> </ul>	Yes
snapshot	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• container (Boolean)</li> <li>• crawlId (Int64)</li> <li>• signature (String)</li> <li>• timestamp (Int64)</li> <li>• error (Boolean)</li> <li>• notFoundCount (int32) *</li> </ul>	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• id (String)</li> <li>• url (String)</li> <li>• fetchUrl (String)</li> <li>• itemType (String)</li> <li>• displayUrl (String)</li> <li>• container (Boolean)</li> <li>• crawlId (Int64)</li> <li>• signature (String)</li> <li>• timestamp (String)</li> <li>• error (Boolean)</li> <li>• notFoundCount (int32)</li> </ul>	No

statistics	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• statistics (Object) <ul style="list-style-type: none"> <li>◦ .@processor (String)</li> <li>◦ .@server (String)</li> <li>◦ .@status (String)</li> <li>◦ .@mode (String)</li> <li>◦ .@startTime (Int64)</li> <li>◦ .@endTime (Int64)</li> <li>◦ .@currentTime (Int64)</li> <li>◦ .@cs (String)</li> <li>◦ .queue (Object) <ul style="list-style-type: none"> <li>▪ .scan (Object) <ul style="list-style-type: none"> <li>• .@toScan (Int32)</li> <li>• .@scanning (Int32)</li> <li>• .@scanned (Int32)</li> <li>• .@total (Int32)</li> </ul> </li> <li>▪ .process (Object) <ul style="list-style-type: none"> <li>• .@toProcess (Int32)</li> <li>• .@processing (Int32)</li> <li>• .@processed (Int32)</li> <li>• .@total (Int32)</li> </ul> </li> </ul> </li> <li>◦ .inProgress (Object) <ul style="list-style-type: none"> <li>▪ .@adding (Int32)</li> <li>▪ .@updating (Int32)</li> <li>▪ .@deleting (Int32)</li> <li>▪ .@total (Int32)</li> </ul> </li> <li>◦ .processed (Object) <ul style="list-style-type: none"> <li>▪ .@added (Int32)</li> <li>▪ .@updated (Int32)</li> <li>▪ .@deleting (Int32)</li> <li>▪ .@unchanged (Int32)</li> <li>▪ .@excluded (Int32)</li> <li>▪ .@terminated (Int32)</li> <li>▪ .@errored (Int32)</li> <li>▪ .@total (Int32)</li> </ul> </li> <li>◦ .errors (Object) <ul style="list-style-type: none"> <li>▪ .@batch (Int32)</li> <li>▪ .@scan (Int32)</li> <li>▪ .@document (Int32)</li> <li>▪ .@total (Int32)</li> </ul> </li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• @processor (String)</li> <li>• @server (String)</li> <li>• @status (String)</li> <li>• @mode (String)</li> <li>• @startTime (Int64)</li> <li>• @endTime (Int64)</li> <li>• @currentTime (Int64)</li> <li>• @cs (String)</li> <li>• queue (Object) <ul style="list-style-type: none"> <li>◦ .scan (Object) <ul style="list-style-type: none"> <li>▪ .@toScan (Int32)</li> <li>▪ .@scanning (Int32)</li> <li>▪ .@scanned (Int32)</li> <li>▪ .@total (Int32)</li> </ul> </li> <li>◦ .process (Object) <ul style="list-style-type: none"> <li>▪ .@toProcess (Int32)</li> <li>▪ .@processing (Int32)</li> <li>▪ .@processed (Int32)</li> <li>▪ .@total (Int32)</li> </ul> </li> </ul> </li> <li>• inProgress (Object) <ul style="list-style-type: none"> <li>◦ .@adding (Int32)</li> <li>◦ .@updating (Int32)</li> <li>◦ .@deleting (Int32)</li> <li>◦ .@total (Int32)</li> </ul> </li> <li>• processed (Object) <ul style="list-style-type: none"> <li>◦ .@added (Int32)</li> <li>◦ .@updated (Int32)</li> <li>◦ .@deleting (Int32)</li> <li>◦ .@unchanged (Int32)</li> <li>◦ .@excluded (Int32)</li> <li>◦ .@terminated (Int32)</li> <li>◦ .@errored (Int32)</li> <li>◦ .@total (Int32)</li> </ul> </li> <li>• errors (Object) <ul style="list-style-type: none"> <li>◦ .@batch (Int32)</li> <li>◦ .@scan (Int32)</li> <li>◦ .@document (Int32)</li> <li>◦ .@total (Int32)</li> </ul> </li> </ul>	No
status	<ul style="list-style-type: none"> <li>• _id (ObjectId)</li> <li>• connectorSource (Object)</li> <li>• @action (String)</li> <li>• @actionProperties (String)</li> <li>• @crawlId (String)</li> <li>• @normalizedCSName (String)</li> <li>• displayName (String)</li> <li>• @scheduler (String)</li> <li>• @scheduleId (String)</li> <li>• @jobNumber (String)</li> <li>• @sourceId (String)</li> <li>• @actionType (String)</li> <li>• @dbId (String)</li> <li>• crawlStart (Int64)</li> <li>• crawlStatus (String)</li> <li>• processDeletes (String)</li> <li>• processingDeletesStatus (String)</li> <li>• crawlEnd (Int64)</li> </ul>	<ul style="list-style-type: none"> <li>• _id (String)</li> <li>• connectorSource (Object)</li> <li>• @action (String)</li> <li>• @actionProperties (String)</li> <li>• @crawlId (String)</li> <li>• @normalizedCSName (String)</li> <li>• displayName (String)</li> <li>• @scheduler (String)</li> <li>• @scheduleId (String)</li> <li>• @jobNumber (String)</li> <li>• @sourceId (String)</li> <li>• @actionType (String)</li> <li>• @dbId (String)</li> <li>• crawlStart (Int64)</li> <li>• crawlStatus (String)</li> <li>• processDeletes (String)</li> <li>• processingDeletesStatus (String)</li> <li>• crawlEnd (Int64)</li> </ul>	No

\*These fields were available in Aspire 3.1