

Hierarchy Extractor

The Hierarchy Extractor looks for the 'hierarchy' tag in a job, and when located, sends jobs to index any new parents, their fields and ACLs.

- [Configuration](#)
 - [Branch Handler Configuration](#)
- [Example Configurations](#)
 - [Simple](#)
 - [Fixed ACLs Configuration](#)
- [Example Output](#)
- [Parent Database Management](#)

Hierarchy Extractor	
Factory Name	com.searchtechnologies.aspire:aspire-hierarchy-extractor
subType	default
Inputs	AspireObject with a 'hierarchy' tag
Outputs	Send jobs to index any new parents, their fields and ACL.

[Aspire Enterprise](#)

Configuration

Element	Type	Default	Description
acls/acl/@usergroup	string		The user/group name for the ACL.
acls/acl/type	string	<i>Allow</i>	Indicates whether the user/group will have access to the crawled files. Options include: <i>allow</i> , <i>deny</i> .
acls/acl/entity	string	<i>group</i>	Specifies if the ACL corresponds to a group or user. Options include: <i>group</i> , <i>user</i> .

If no fixed ACLs configured as above, then a union of parent plus children ACLs is going to be used as the ParentACLs, and each time a new child adds a new ACL to the Union, the parent job is going to be reindexed.

Branch Handler Configuration

This component publishes to the *onAdd*, *onDelete* and *onUpdate*, so a branch must be configured for each of these three events.

Element	Type	Description
branches/branch/@event	string	The event to configure - <i>onAdd</i> , <i>onDelete</i> or <i>onUpdate</i> .
branches/branch/@pipelineManager	string	The name of the pipeline manager to publish to. Can be relative.
branches/branch/@pipeline	string	The name of the pipeline to publish to. If missing, publishes to the default pipeline for the pipeline manager.
branches/branch/@allowRemote	boolean	Indicates if this pipeline can be found on remote servers (see Distributed Processing for details).
branches/branch/@batching	boolean	Indicates if the jobs processed by this pipeline should be marked for batch processing (useful for publishers or other components that support batch processing).
branches/branch/@batchSize	int	The max size of the batches that the branch handler will created.
branches/branch/@batchTimeout	long	Time to wait before the batch is closed if the batchSize hasn't been reached.
branches/branch/@simultaneousBatches	int	The max number of simultaneous batches that will be handled by the branch handler.

Example Configurations

Simple

```
<component name="HierarchyExtractor" factoryName="aspire-hierarchy-extractor" subType="default">
  <branches>
    <branch event="onAdd" pipelineManager="." pipeline="addPipeline" batching="true"/>
    <branch event="onDelete" pipelineManager="." pipeline="deletePipeline" batching="true"/>
  </branches>
</component>
```

Fixed ACLs Configuration

```

<component name="HierarchyExtractor" factoryName="aspire-hierarchy-extractor" subType="default">
  <acls>
    <acl usergroup="mycompany\aaguilar">
      <type>allow</type>
      <entity>user</entity>
    </acl>
    <acl usergroup="mycompany\stAllEmployees">
      <type>deny</type>
      <entity>group</entity>
    </acl>
  </acls>
  <branches>
    <branch event="onAdd" pipelineManager="." pipeline="addPipeline" batching="true"/>
    <branch event="onDelete" pipelineManager="." pipeline="deletePipeline" batching="true"/>
  </branches>
</component>

```

Example Output

For every new parent found a job will be sent to the "onAdd" event of the branch handler:

```

<doc source="/HierarchyExtractor/Main/HierarchyExtractor">
  <hierarchy>
    <item id="CDCE0D45AC20FDE62F5CEB6118643033" level="1" name="FSC" type="aspire/filesystem" url="C:\testdata\a\">
      <ancestors/>
    </item>
  </hierarchy>
  <id>C:\testdata\a\</id>
  <url>C:\testdata\a\</url>
  <fetchUrl>C:\testdata\a\</fetchUrl>
  <action>add</action>
  <md5>CDCE0D45AC20FDE62F5CEB6118643033</md5>
  <mimeType>aspire/filesystem</mimeType>
  <lastModified>2014-03-21T17:44:20Z</lastModified>
  <dataSize>0</dataSize>
  <content>url:C:\testdata\a\ docId:CDCE0D45AC20FDE62F5CEB6118643033</content>
  <sourceName>FSC</sourceName>
  <sourceType>filesystem</sourceType>
  <acls>
    <acl access="allow" domain="mycompany" entity="user" fullname="mycompany\aaguilar" name="aaguilar" scope="global"/>
    <acl access="deny" domain="mycompany" entity="group" fullname="mycompany\stAllEmployees" name="stAllEmployees" scope="global"/>
  </acls>
</doc>

```

Parent Database Management

There are 5 servlet commands you can use to manage the parent database, available from the debug console:

- Reindex
Resend the jobs to the "onAdd" event of the configured Branch Handler
- Dump
Creates a dump file of the database, that you can import later
- Import
Imports the data from a dump file from the file system
- Clear
Deletes all content from the database, you can decide if you want to send delete jobs to the "onDelete" branch of the configured Branch Handler.
- Statistics
Return the count of parents stored in the database