Aspire for Connectors and Content Processing
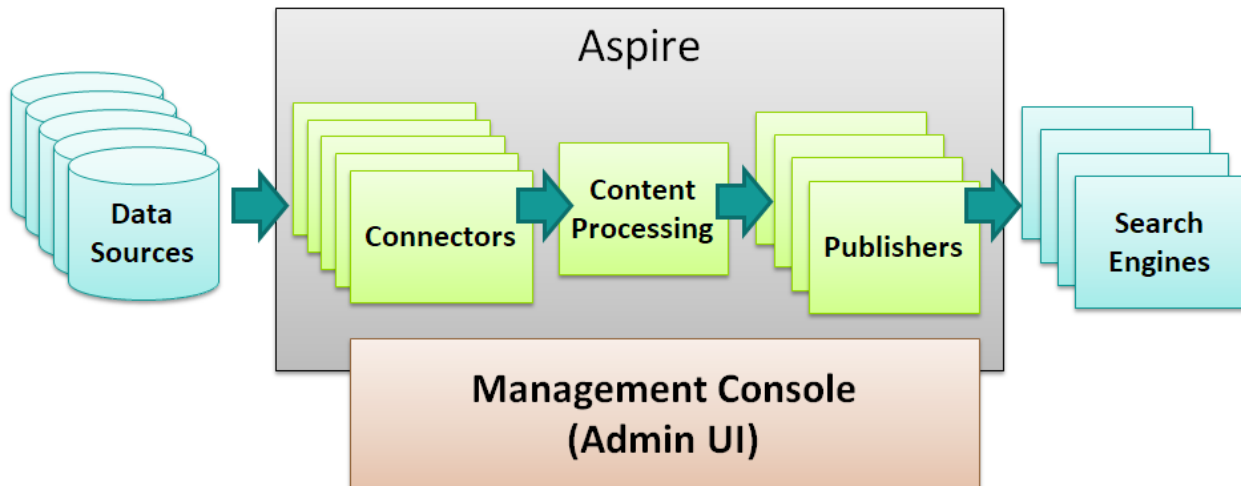
Welcome to Aspire!

Aspire is both a framework and a complete end-to-end content ingestion and content processing system.

Aspire for Connectors and Content Processing

Typically, Aspire is used as an end-to-end system for acquiring content, processing it, and publishing it to be indexed by search engines:

All of this is can be done within a single Aspire node (running on a single JVM) or across a cluster of machines cooperatively working together.

Aspire Features

- Built-in connectors to dozens of different data sources
    - *Scalable*: Automatically distributes ingestion jobs across a cluster of nodes
    - *Elastic*: Add and remove nodes at any time
    - *Resilient*: Crawl state is carefully tracked at all points
        - Jobs on failed nodes are automatically picked up by other nodes
        - After a full system crash, crawling restarts from where it left off
    - *High Performance*: Crawls are typically limited only by limitations on the source system
    - *Incremental*: Automatically identifies incremental changes and processes only those changes
        - The method for detecting incremental changes is based on what is provided by the underlying content storage technology.
- Built-in publishers to most commonly available search engines
    - Including Solr, Elasticsearch, SharePoint, the GSA, and others
- Built-in components for many common content processing tasks
    - Such as text extraction, OCR, field mapping, domain mapping, etc.
- Scripting for easy manipulation of metadata
- Fully understands document-level security
    - Ingests ACLs for each content source
    - Provides cached, high-performance group-expansion for each content source
- Extensible
    - Create custom connectors and publishers
    - Create custom pipelines and workflow controls
    - Create custom components
- Ease of deployment
    - Components and configurations are deployed through Maven
    - Properties allow for anything to be parameterized (e.g. server destinations, file directory locations, etc)
    - Content source configurations can be exported from any cluster and imported on another

Product Categories

Note that Aspire components are only available for customers who purchase a connector license. See Aspire Product Categories for more information.

Where to go from here

If you want to use the Aspire strictly as a component and pipeline processing machine, we recommend you use the framework.

If you want to use the connectors and publishers, we recommend that you run the Getting Started Tutorial.