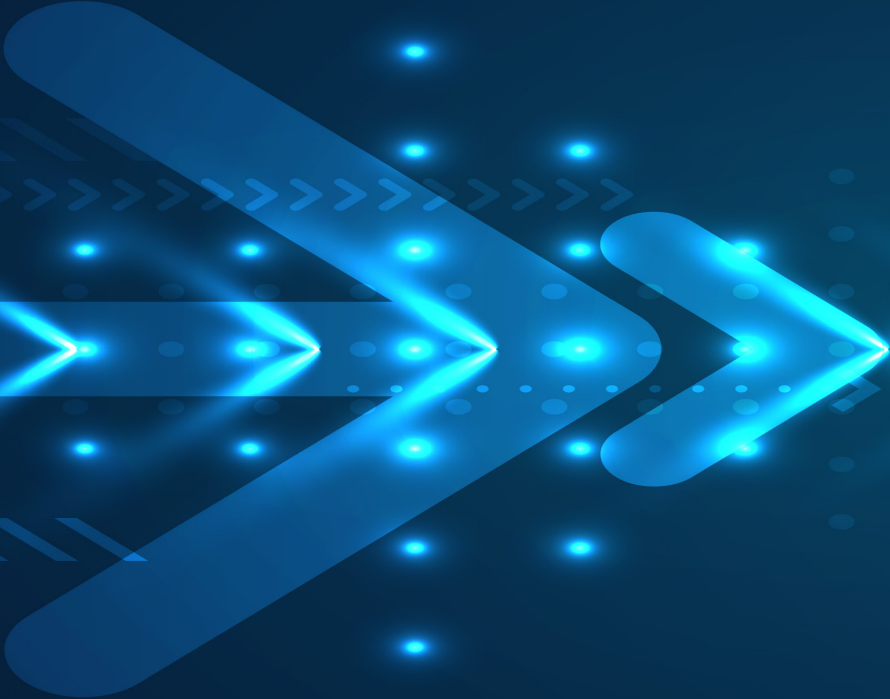


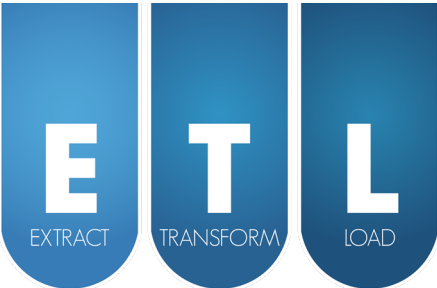
INTRODUCING

LUCILLE



AN OPEN-SOURCE, PRODUCTION-GRADE SEARCH ETL SOLUTION

THE BEST PRACTICES IN SEARCH ETL CODIFIED INTO A SINGLE SOLUTION.



WHAT IS SEARCH ETL?

Search ETL is a category of ETL solutions where data must be extracted from a source system, transformed, and loaded into a search engine – either in batches or in a streaming fashion.

A SEARCH ETL SOLUTION MUST SPEAK THE LANGUAGE OF SEARCH.

It must represent data in the form of search-engine-ready Documents.

It must know how to enrich Documents to support common search use cases.

It must follow best practices for interacting with search engines including support for bulk updates, routing, and versioning.

A PRODUCTION-GRADE SEARCH ETL SOLUTION SHOULD BE SCALABLE, RELIABLE, & EASY TO USE.

It should support parallel Document processing.

It should be observable and auditable.

It should be easy to configure.

It should have extensive test coverage, and it should have been hardened through multiple challenging real-world deployments.

[kmwtechnology/lucille](https://github.com/kmwtechnology/lucille)

THIS IS LUCILLE: A PRODUCTION-GRADE SEARCH ETL SOLUTION

LUCILLE CODIFIES

SEARCH ETL

BEST PRACTICES.

SO YOU

DON'T HAVE TO.

Lucille helps you get your data into Lucene-based search engines like **OpenSearch**, **Apache Solr**, and **Elasticsearch**, as well as vector-based search engines like **Pinecone** and **Weaviate**.

It helps you keep that search engine content up-to-date as your backend data changes. Lucille does this in a way that scales as your data volume grows, and in a way that's easy to customize as your data transformation requirements become more complex.

Lucille implements search best practices so you can stay focused on your data itself and what you want to do with it.

GET STARTED WITH LUCILLE TODAY

Lucille is an open source Search ETL solution. Simply clone the repo and get started by looking at some example modules or by following a simple practice exercise in Lucille's README.

Interested in a managed Lucille service? Contact KMW Technology and we will be glad to help get you started.

Created by Search Experts focused on resolving pain points

Lucille is an open source Search ETL solution created by people who work in search every day and who know the traditional pain points associated with getting your data processed, enriched and indexed into a search engine.

Lucille is designed to reduce or eliminate these pain points.

Easy to stand up, yet massively scalable.

It is easy to stand up for local experimentation and POCs while remaining massively scalable for high throughput, low latency production environments.

Flexible data connectors

It is flexible with the ability to connect and extract data from databases, file systems, CSVs, JSON, XML, Parquet and more.

Pluggable stages

Lucille offers a wide variety of document processing pipeline stages including text extraction using Apache Tika, dictionary-based entity extraction, and support for text extraction via OCR coming soon

Easily configurable

Lucille is configured using a simple configuration file.

Highly performant

Lucille is designed to be highly performant, often indexing documents at a rate of thousands of documents per second.



github.com/kmwtechnology
info@kmwllc.com
kmwllc.com

SEARCH & AI PROFESSIONAL SERVICES

As regular contributors to Open Source code bases, the KMW team supports the ecosystem of search developers worldwide. Lucille is KMW's latest contribution to the open source world and reflects lessons and expertise KMW has acquired from creating varied search solutions for organizations small and large.