Lucidworks

LUCIDWORKS® ETL INDEXER

Fast bulk indexing that recovers on its own.

Load large volumes of structured data quickly, with automatic recovery and full visibility into every run.

What it is

The Lucidworks ETL Indexer is a bulk indexing service for Lucidworks Search™. It loads large volumes of structured data from cloud storage into your search collections, quickly and reliably.

It runs continuously, picks up each indexing request as soon as it arrives, and carries it through to completion. If a run is interrupted, it starts that request again on its own.

The problem

Bulk loading often relies on batch jobs that are fragile and hard to see into:

- **Jobs that stop on error** and have to be diagnosed and restarted by hand
- **Little visibility** into what was indexed, what failed, and why
- **Support tickets** for problems a team could fix itself with the right view

The result: stale search results, partial indexes, and slow recovery.

How it solves it

The ETL Indexer is an always-on service built for speed and recovery:

- **Parallel indexing** that reads files in large batches and loads collections side by side
- **Automatic restart** of any request that was interrupted mid-run
- **A visible lifecycle** for every request: pending, processing, completed, or failed
- **Per-collection metrics** for documents indexed and documents that failed

Why it matters

- **Fresher results.** New data is indexed as soon as a request lands.
- **Fewer support tickets.** Teams can see failed runs and recover them on their own.
- **More complete indexes.** Interrupted runs start again instead of stopping halfway.
- **Faster recovery.** Re-index from data already in cloud storage, with no re-upload.

`Always on: requests start as soon as they land, and an interrupted run starts again by itself.`

How it works

1. Submit a request from the Jobs screen, or index directly to Solr.
2. The indexer picks it up and reads your files in large batches.
3. Documents go straight to Solr, or through a Lucidworks index pipeline.
4. Collections load in parallel or in sequence, by indexing group.
5. The request moves to completed, or to failed with the reason logged.

Key benefits

01

Speed at scale

Parallel indexing moves large files fast, tuned to your hardware.

02

Resilient by design

Interrupted runs restart on their own. Every failure logs a reason.

03

Visible without a ticket

Per-collection metrics and a clear lifecycle show how each run went.

Technical details

Data formats

Parquet, CSV, JSON, and JSONL

Data source

Google Cloud Storage: one file, or a pattern matching many

Destinations

Solr directly, or a Lucidworks index pipeline

Transformations

Copy, rename, delete, and add fields by configuration

Monitoring

Per-collection metrics in Prometheus and Grafana

Availability

Lucidworks Search 5.19

Make bulk indexing the easy part.

Talk to your Lucidworks account team about turning on the ETL Indexer for your Lucidworks Search environment.

Lucidworks

www.lucidworks.com

© 2026 Lucidworks, Inc.
Lucidworks is a registered trademark.