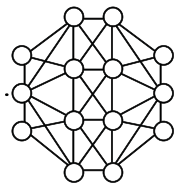
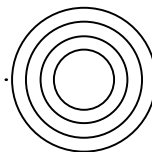


The **ETL+** Platform for GenAI.

[illegible]

LLM

Top 100 AI Companies

ETL **Plus+** so much more.

UNSTRUCTURED

Extract

Source Connectors

Airtable *
Astra DB *
Azure Blob Storage
Box
Confluence
Couchbase
Databricks Volumes
Discord *
Dropbox
Elasticsearch
GitHub *
GitLab *
Google Cloud Storage
Google Drive
Jira
Kafka *
Local File System
MongoDB
Notion *
OneDrive
OpenSearch *
Outlook
PostgreSQL
S3
Salesforce
SFTP *
SharePoint
SingleStore *
Slack *
Snowflake
SQLite *
VastDB *
Zendesk

* Preview Connectors

FILE

Transform

File Types

.abw	.docx	.gif	.mcw	.png	.rtf	.uof	.xlsm
.bmp	.dot	.heic	.mw	.pot	.sdp	.uos1	.xlsx
.csv	.dotm	.htm	.odt	.potm	.sgl	.uos2	.xlw
.cwk	.eml	.html	.org	.ppt	.svg	.web	.xml
.dbf	.epub	.hwp	.p7s	.pptm	.sxd	.webp	.zabw
.dif	.et	.jpeg	.pages	.pptx	.tiff	.wk2	
.doc	.eth	.jpg	.pbd	.prn	.txt	.xls	
.docm	.fods	.md	.pdf	.rst	.tsv	.xlsb	+more

Partitioning (RAW to Structured JSON)

Fast	Rules-based parser for documents with extractable text
Hi-Res	High speed CPU enabled object detection and OCR pipeline
VLM	Complex diagrams, tables, and image descriptions
Video-to-text	Custom frame-by-frame summarization
Speech-to-text	Rich transcripts from audio or video recordings

✓ Amazon Bedrock	✓ Google Vertex AI
✓ Anthropic Claude	✓ OpenAI GPT-4o
✓ Google Gemini	✓ NVIDIA NeMo Retriever + VLM

Chunking

✓ Character	✓ Page
✓ Contextual	✓ Similarity
✓ Custom	✓ Title

Enrichments

- ✓ NER (Name entity recognition), Topics, and Triples
- ✓ Image and table summarization
- ✓ Metadata tags (over 45 and counting)

Embeddings

✓ Amazon Bedrock	✓ Voyage AI
✓ Azure OpenAI	✓ NVIDIA NeMo Retriever
✓ Together AI	Text Embedding

Load

Destination Connectors

Astra DB
Azure AI Search
Azure Blob Storage *
Chroma *
Couchbase
Databricks Volumes
Databricks Delta Table
AWS Delta Tables
DuckDB
Elasticsearch
Google GCS
IBM Milvus
IBM watsonx.data
Zilliz
Kafka
KDB.ai *
Lance DB *
MongoDB
Neo4j
OneDrive
OpenSearch *
Pinecone
PostgreSQL
Qdrant
Redis
S3
SingleStore *
Snowflake
VastDB *
Vectara *
Weaviate

* Preview Connectors

JSON

Chunked
Embedded
Metadata

Plus+

Automation

- ✓ Scheduled Jobs
- ✓ Smart Document Routing
- ✓ Auto Workflow Optimization
- ✓ Intelligent Synchronization

Interfaces

- ✓ API
- ✓ UI
- ✓ MCP

Connectors & Plugins

- ✓ 24/7 Connector Maintenance
- ✓ Universal Connectivity
- ✓ Secure Credential Handling
- ✓ Modular Plugin Architecture

Enterprise-Ready

- ✓ Company + Team Accounts
- ✓ Role-Based Access Control
- ✓ Robust Error Handling & Observability
- ✓ HIPAA, SOC2 Type 2, GDPR, + more

Deployment

- ✓ SaaS
- ✓ Hybrid SaaS
- ✓ VPC
- ✓ Bare Metal

Schedule a Demo

sales@unstructured.io
unstructured.io

