

ReportMiner 8

Convert Raw Unstructured Data into Intelligent Insights

Cut Down Data Preparation Time from Weeks to Hours

Accelerate data retrieval, cleaning, and reformatting to prepare unstructured data for analysis using the self-service approach.

Create Reusable Data Extraction Models for Repeatable Processes

Build flexible, template-based data extraction models for files with similar layouts and re-use them for faster process execution.

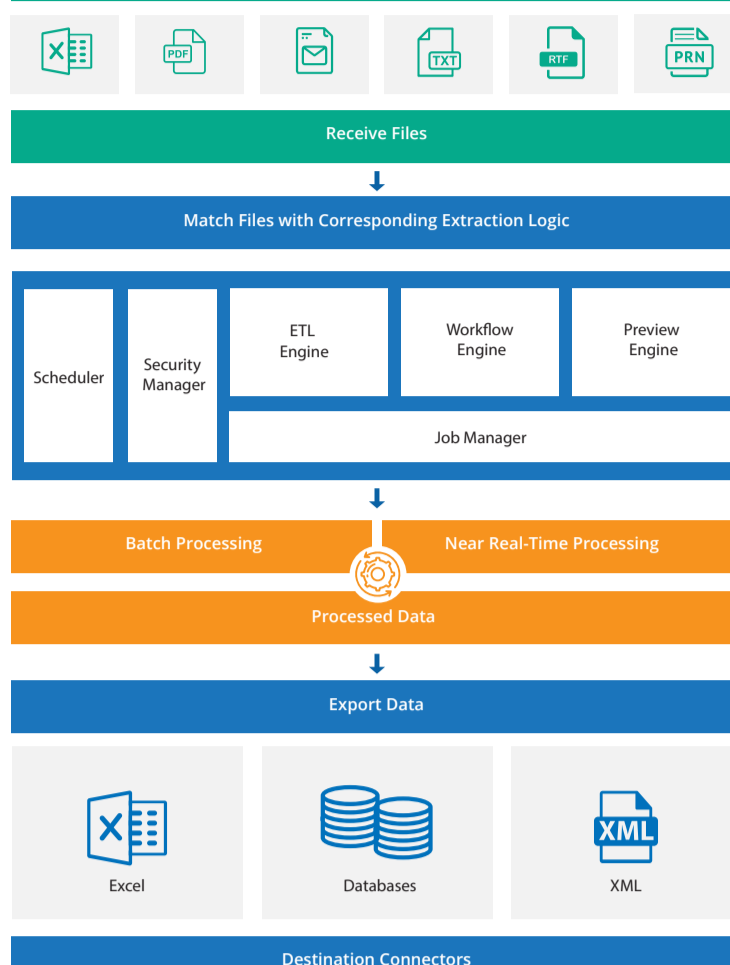
Integrate Extracted Data Using an Intuitive, Visual Approach

Create integration processes using drag-and-drop mapping and code-free UI to reduce the gap between data preparation and analysis.

Fast information access is crucial to the successful and efficient implementation of analytical processes. When you're dealing with unstructured data trapped within bank statements, invoices, spreadsheets, machine logs, and reports - retrieving data manually requires a significant investment of time and resources. The result is delayed insights and inefficient decision-making.

Astera's self-service data preparation platform, ReportMiner, enables business users and analysts to extract, transform, clean, standardize, and convert data into their desired format - without writing a single line of code. The combination of intuitive, visual UI and process automation features enables users to create repeatable data preparation processes yielding substantial cost savings in the long run.

ReportMiner is purpose-built to support the extraction and structuring of unstructured data sources to be used within any data pipeline and solve complex operational problems for faster time-to-insights.



Astera ReportMiner Architecture

Development Environment

Intuitive UI: Prepare your raw, unstructured enterprise data for further processing in a visual, easy-to-use interface.

Dedicated Properties Panel	View and modify report model properties, and observe the changes in real-time through dedicated panels for Report Options, Region, Field, and Pattern properties.
Shortcut Icons	Specify Report, Region, Field, and Pattern properties through the shortcut icons present in the Properties Panel toolbar, eliminating the need to switch screens.
Toolbox	Access data connectors and pre-built transformations through the Toolbox to fuel your data preparation processes.
Report Browser	Manage your report models with a Report Browser that shows the complete model layout, and provides easy access to data import and export options.
Single-Click Preview	Verify and test the report model as it's being built or modified via Single-Click Preview. View the extracted data in real-time and check for verification errors.
Project Explorer	Get access to all the elements of your data extraction projects, including report models and shared connections, in one place with the Project Explorer.

Data Extraction: Retrieve data from a range of unstructured and semi-structured sources using built-in source connectors.

File Sources	Capture meaningful information from diverse file sources, including PDFs, PDF forms, TXT, PRN, RTF, XLS, XLSX, and COBOL.
File Repositories	Leverage native connectivity to file repositories, such as Local Directory, FTP/SFTP, and e-mail source to expedite data retrieval.
Hierarchical Data	The hierarchical model capability of ReportMiner enables extraction from complex hierarchical text reports.
REST APIs (available as an add-on)	Access data from cloud sources using internal, external, and custom APIs by importing JSON/YML file or URL.

Data Export: Transfer extracted data to a range of destinations and formats.

Databases	Load data in your required database system. Supported destinations include relational databases, like SQL Server, MariaDB, SAP HANA, Vertica, Redshift, Aurora, Snowflake, Oracle, IBM DB2, MySQL, PostgreSQL, and others via ODBC connectivity.
File Formats	Export data to different file formats, including fixed length, .CSV, Excel, and XML/JSON.
Export as Dataflow	Use your report model as a source object in a dataflow, and leverage the advanced pre-built transformations and SQL Query features of ReportMiner to create advanced data processes.
Export Settings	Modify export settings at the final stage by filtering data, adding expressions, and more without making any changes in the report model.

Report Model

Data Extraction Models: Build reusable extraction models in an easy-to-use interface and cut down manual efforts in your data preparation process.

Create New Model	Build new report models using the simple, point-and-click UI. View all the report model's occurrences in the Report Browser.
Re-Use Model	Save and use extraction models for subsequent conversions involving files with similar layout, saving time and manual effort.
View Statistics	Get detailed statistical breakdown of the extracted data, including error count, null count, warning count, and more to ensure data accuracy and validity.
Pattern	Capture unstructured data into a structured format by specifying a pattern based on an alphabet, character, word, number or a combination of these.
Data Region	Specify the number of lines covered in a source report to create a data region that is specified by a pattern.
Data Field	Extract data points from the file and write them as elements of a data field within the data region.
Match Entities	Identify the starting point of a region by matching an alphabet, character or digit, and creating a text pattern.

Report Properties

Configure Reports: Modify the extraction properties without switching windows.

Reading Options	Identify the starting point of a region by matching an alphabet, character or digit, and creating a text pattern.
-----------------	---

Define Regions: Create report models by defining region properties in the sidebar panel.

Auto Create Fields	Specify which data to extract from your unstructured reports by creating your data fields with just one-click.
Auto Determine Field Names	Eliminate manual efforts by automatically determining the field names in the selected data region for new source files loaded into the report model.
Region Details	Customize your data regions by specifying where you want to end your region. The available options include <i>Line Count</i> , <i>Blank Line</i> , <i>Last Field Ends</i> , and <i>Another Region Starts</i> .
Specify Container	Specify the constraints of your data region by enclosing it in a container or defining overlapping containers.

Create Fields: Support the selection and extraction within a data field through the set of field properties displayed in the sidebar.

Auto-Determine Field Length	Determine the field length of the available data with a single click, cutting down manual efforts.
Modify Field Length	Easily manage the field length by increasing or decreasing input length by the desired number of characters.
Data Cleansing	Utilize the enhanced built-in data cleaning options to remove redundant characters and spaces from report models.
Modify Size and Position	Manage misalignments by precisely identifying the field's start position, and modifying size and length properties.

Adjust Patterns: Create patterns in the report model to search for relevant data.

Specify Pattern Count	Capture a data region based on the pattern count specified in the properties panel to identify the starting point of your region.
Modify Pattern Properties	Match the pattern with the data based on Case Sensitive Pattern Match, Regular Expression, or Floating Pattern.
Multi-Column	Split a region into multiple columns and visually define the layout of the column dividers by selecting the Multi-Column checkbox.

Workflow Automation

Process Orchestration: Execute a series of repeatable data extraction jobs using workflows and dataflows.

Run Jobs	Execute dataflows, workflows, as well as transfer and batch jobs on a single server or distribute them across multiple servers.
Job Scheduling	Specify triggers based on file drops or running frequency using the built-in Job Scheduler to automate repeatable processes.
Send Email	Track the workflow progress by setting up alert emails for abnormal terminations, task completions, errors, and other dataflow operations.

Performance

Performance Optimization: Leverage ReportMiner's performance-driven features for fast execution of your data preparation processes.

Job Optimizer	Remove unnecessary sort operations and ORDERBY clauses to reduce job execution time by utilizing the built-in Job Optimizer.
Near Real-Time Processing	Automatically pull files in near real-time using the File Drop feature and orchestrate the entire data extraction process with minimal manual efforts.
Bulk File Processing	Run multiple reports with one report model by looping through a folder with your documents and ensure faster processing.

Security Controls

User Authorization: Manage user access via advanced authentication protocols.

Bearer Token Authentication	Generate a username, password, and Token URL to regulate access requests to the server.
Secure Domain Communication	Encrypt information transmission between server and client using the HTTPS commands.

User Authentication: Create user-defined roles to control and manage user access.

Role Management	Control role-based access for new users and modify the existing ones via the role management feature.
User Management	Manage users and their roles through a unified user management interface.

Monitoring

Monitoring Protocols: Track the progress of your extraction jobs and monitor the server load via monitoring features.

Job Monitoring	Get a detailed summary of the data extraction jobs executed on the server with detailed information about each step to control the re-execution or termination of dataflows and workflows.
Server Monitoring	Generate a <i>Server Diagnostics File</i> and <i>Log File</i> to manage server resource utilization and enable faster troubleshooting.