MACHINE DATA INSIGHTS

CIM Normalization Pipeline

Turn raw machine data into validated, deploy-ready content for Splunk® ES & ITSI – faster, with AI-accelerated tooling and repeatable methods.

There's Gold In That Data!®

THE FOUNDATION

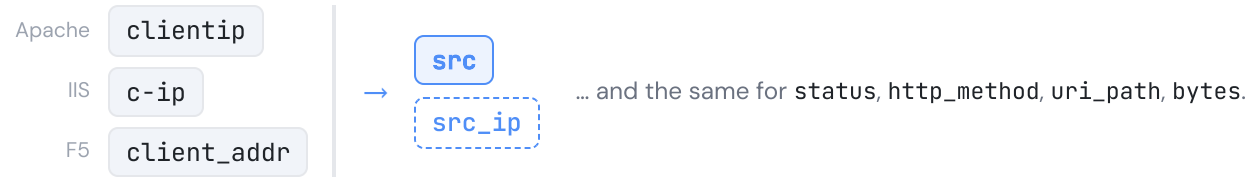
What is CIM Normalization?

CIM normalization aligns security and operational data from every vendor to Splunk's standard schemas – so **one detection, dashboard, or KPI works across all of them**, instead of a custom search per technology.

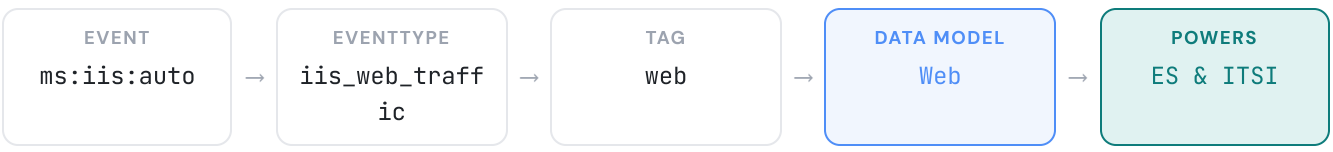
- **Map** vendor-specific fields to common CIM fields.
- Create **eventtypes** that identify significant security and operational events.
- Create **tags** that link those events to the accelerated data models Splunk ES & ITSI run on.
- Make every source **analyzable consistently** – firewall, endpoint, identity, cloud, email, PLC, SCADA, IoT.

From Vendor Fields to Detections

1 · DISPARATE VENDOR FIELDS MAP TO COMMON CIM FIELDS



2 · EVENTTYPES AND TAGS CONNECT THOSE EVENTS TO THE DATA MODELS



WHY IT MATTERS

What CIM Normalization Unlocks in Splunk ES & ITSI

You can write detections without CIM - most teams have, one search per technology. Normalized data is what lets a **single detection, KPI, or dashboard span every vendor**, run at data-model speed, and use the content Splunk ships. ITSI KPIs and entity rules want the same thing: **consistent fields and identifiers**. Miss it and nothing errors - you just pay for coverage one search at a time.

FOR SPLUNK ES

- One correlation search covers **every vendor** in a data model - not one per technology.
- Out-of-the-box **Security Content** works as shipped.
- Better detection accuracy, **fewer false positives**, faster investigations.
- Asset & identity enrichment keys on **CIM fields** - it works even where detections don't use data models.

FOR SPLUNK ITSI

- One **KPI** definition measures every vendor's tier the same way.
- Consistent identifiers mean **entity rules match** - hosts don't drop out of services.
- **Content packs** depend on well-built add-ons; the pipeline produces them.
- Service health scores you can **actually trust** - and explain to the business.

WHAT BREAKS WHEN YOU MOVE A DETECTION ONTO A DATA MODEL WITHOUT CIM NORMALIZATION

Converting a search to run on a data model is where unmapped sourcetypes surface. The data is in Splunk. The search runs clean. It finds **nothing** - because those events never reached the model it queries. **MDI's CIM Assessment Toolkit (CAT) finds them first, ranked by impact.**

The MDI CIM Normalization Pipeline

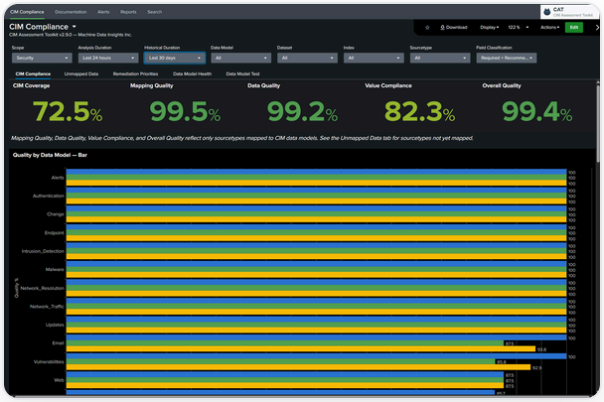
You run the client-side tools; the encrypted exchange hands data to MDI; MDI runs the engine.

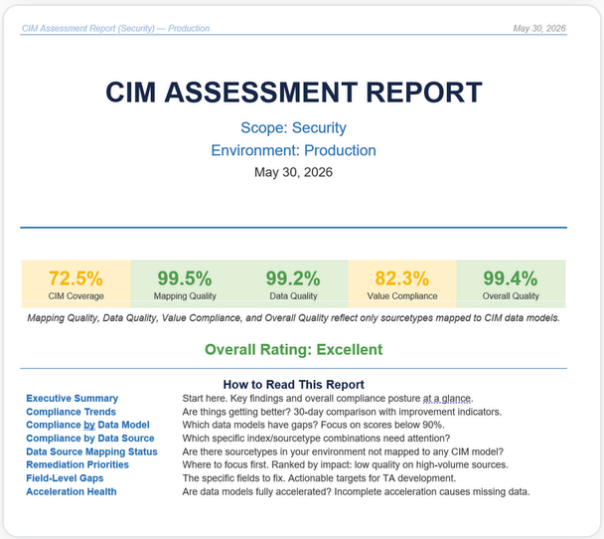
01	→	02	→	03	→	04
Assess		**Sanitize**		**Secure Handoff**		**Build**
CAT		Paydirt		Secure Exchange		Data Refinery

CAT

CIM Assessment Toolkit · Measure What ES Actually Sees

CAT is a free, open-source Splunkbase app that measures CIM (Common Information Model) compliance at the **dataset and sourcetype level** - not just the data model level - so you find the gaps that silently break Splunk ES correlation searches. Five KPIs, an impact-ranked remediation queue, an 850+ sourcetype inventory reference, and an automated executive Word report. One-time setup - a summary index and a few macros - then CAT scores your whole environment automatically. Splunk Enterprise 9.0+ with Splunk_SA_CIM.





Get CAT on Splunkbase →

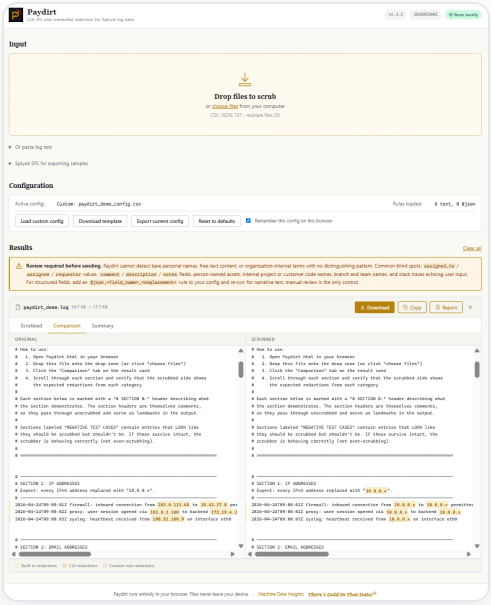
Scan for the Splunkbase listing

Free & Open Source · Splunkbase Published · Dataset-Level Assessment · Acceleration Health · Data Model Testing · Unmapped Sourcetypes

Paydirt

Log Scrubber · There's Gold in That Data!®

Paydirt is a free, open-source tool that scrubs CUI, PII, PHI, and credentials from Splunk or other log data exports so they can be safely shared and analyzed. It runs entirely on your own machine - as a self-contained browser tool or a Python CLI - with no installation and no network calls. Drop a file, get a sanitized version back. CMMC, HIPAA, and GDPR aware.



github.com/machinedatainsights/paydirt →

Free & Open Source | Runs Offline | No Install

Secure Exchange

Encrypted client data transfer - in and out

Client data never moves by email or thumb drive. Every engagement gets its own private, encrypted exchange: clients drop source data into their inbound folder and collect signed, documented deliverables from their outbound folder - and see nothing else.

IN Client → encrypted upload → MDI pipeline

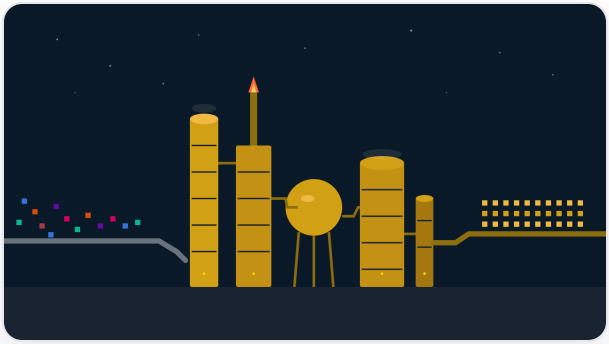
OUT MDI pipeline → signed deliverables → Client

- Per-client isolation - each client sees only their own folder
- Encrypted in transit and at rest
- Immutable audit trail - every upload and download logged
- Automatic notification on every client upload

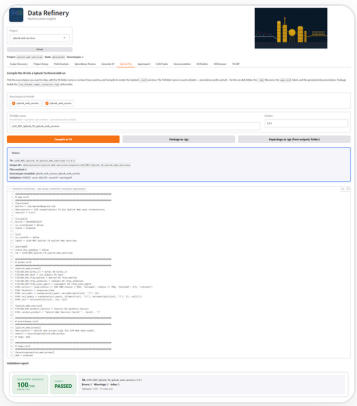
How it's delivered: MDI-operated - we provision and run the exchange; clients upload and download from a browser, with credentials scoped to their own engagement.

Data Refinery

Normalization Pipeline · Raw Exports In, Deploy-Ready Artifacts Out



Turning Data Into Gold™



Splunk TAs tab - compile, validate, package

Data Refinery is the engine MDI uses to turn machine-data exports into validated, CIM-compliant Splunk Technical Add-ons and Cribl packs. It profiles fields and sample events, classifies each sourcetype against the relevant CIM data models, and generates the Splunk `.conf` files (props, tags, etc) that normalize vendor fields to CIM - then compiles, validates, and packages the result.

Where an assessment finds the gaps, Data Refinery builds the fix. Every add-on is linted for CIM compliance, scored for deployment readiness, and run through Splunk AppInspect before it ships - so what you receive is ready to deploy, not a starting point someone still has to finish.

CIM Normalization Summary — Production — Splunk Web Services

Last update: 2026-05-11T18:24:01Z

Project Overview

Technical Add-on

TA Name	Package
zz50_Splunk_TA_splunk_web_services (1.0.0)	zz50_Splunk_TA_splunk_web_services_100.tgz

Indexes and Sourcetypes

Sourcetype	Indexes	Event Count	Description
splunk_web_access	_internal	7,559	HTTP access logs from Splunk Web UI, capturing user requests to the Splunk web interface including static resources, API endpoints, and page loads with client IP, method, status, URI, user agent, and response time.
splunk_web_service	_internal	21,984	Splunk web application server internal performance and diagnostic logs, including page render times, view loading metrics, translation retrieval, RUM (Real User Monitoring) telemetry, and appserver startup information.

Classification and Applicable Data Models

Sourcetype	Classification	Full Name	Security Relevance	CIM Data Models & Datasets
splunk_web_access	—	Application Telemetry	Medium	Web.Web
splunk_web_service	—	Application Performance Monitoring	Low	—

Data Model Macros

Macro Name	Sourcetypes	Expression
cim_Web_indexes	splunk_web_access	(index IN (_internal) AND sourcetype IN ("splunk_web_access"))

Sourcetype: splunk_web_access

Sourcetype	Description
splunk_web_access	HTTP access logs from Splunk Web UI, capturing user requests to the Splunk web interface including static resources, API

Automated CIM Normalization Summary - generated per project

Inside the pipeline

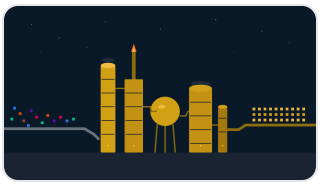
- Scope discovery & sourcetype inventory
- CIM data-model classification
- Splunkbase TA review & coverage scoring
- CIM validation & deployment-readiness scoring
- Packaging (.tgz) & automated Word documentation
- Field-extraction analysis & sample profiling
- KB-backed vendor-to-CIM field mapping
- TA compilation (props / transforms / eventtypes / tags)
- Splunk AppInspect integration (cloud + local)
- Cribl pack generation

How it's delivered: MDI-operated - you receive deployment-ready Splunk TAs and Cribl packs, validated and fully documented.

CIM-Validated Output | AppInspect-Checked | Fully Documented | Splunk TAs + Cribl Packs

One Validated Output, Two Deployment Paths

Data Refinery
CIM-validated output



→

Splunk CIM TAs

props, transforms, eventtypes & tags - normalize vendor fields to CIM at search / index time. Drop-in for Splunk ES.

Cribl Packs

normalize *and* reduce in-stream - cut ingest and license cost before data ever lands.

Let's turn your data into detection.

MDI delivers CIM normalization, data-volume reduction, and CIM macro & correlation-search optimization - with AI-accelerated tooling that cuts time-to-value and consulting cost.

CIM Normalization | Data-Volume Reduction | Correlation-Search Optimization | CIM Macro Optimization | AI-Accelerated

hello@machinedatainsights.com **machinedatainsights.com**

SCAN TO EXPLORE

There's Gold In That Data!®