Guida all'Acquisizione di Dati su Piattaforma Splunk

A cura di Canole Iris

Introduzione

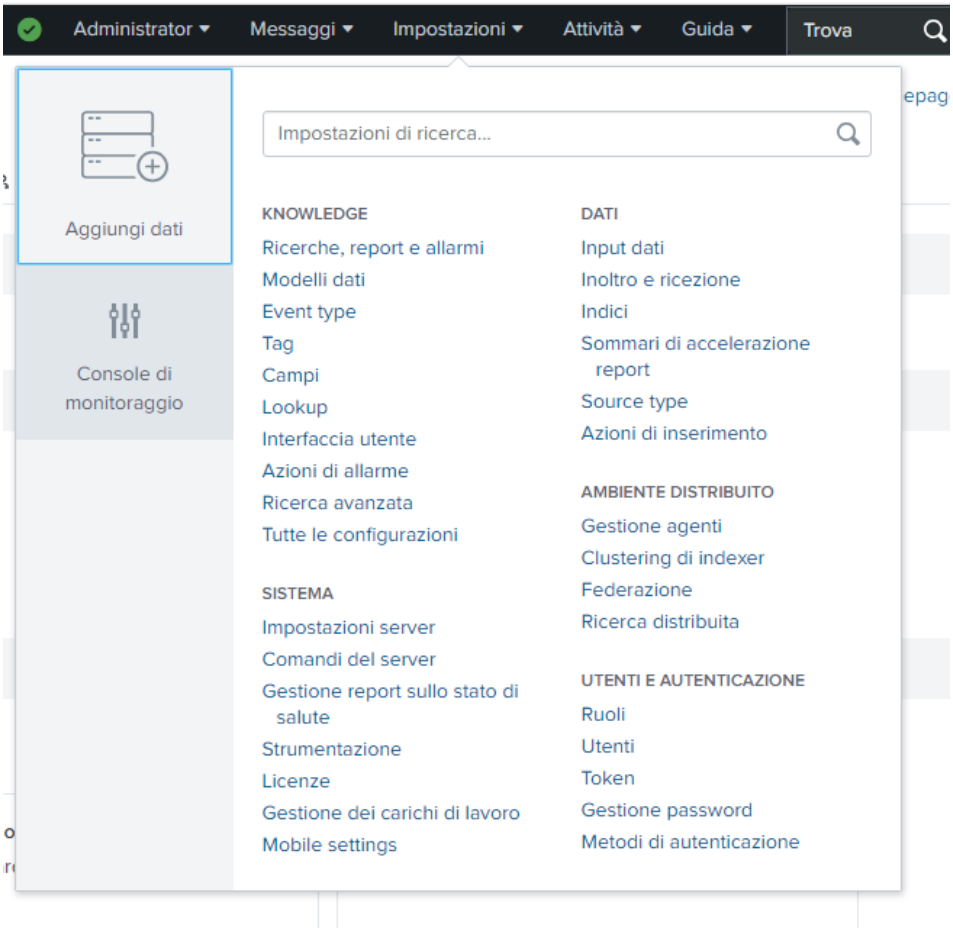
Questo manuale ha lo scopo di illustrare, passo dopo passo, il processo di configurazione di una nuova sorgente dati all'interno della piattaforma Splunk. L'acquisizione dei dati, o *data ingestion*, rappresenta il primo e fondamentale passo per qualsiasi attività di monitoraggio, analisi e reporting. Senza dati affidabili e correttamente indicizzati, le potenti funzionalità di Splunk non possono essere sfruttate appieno.

Nel corso di questa guida, utilizzeremo un esempio pratico e comune: l'acquisizione dei log di eventi generati localmente da un sistema operativo Windows. Seguendo le procedure descritte, l'utente sarà in grado di replicare il processo e acquisire una solida comprensione dei meccanismi di base per l'arricchimento della piattaforma con nuove fonti informative.

1. Accesso alla Sezione di Aggiunta Dati

L'acquisizione di dati è il fondamento su cui poggia l'intero valore della piattaforma Splunk. Pertanto, la funzione "Aggiungi dati" non è semplicemente un pulsante, ma rappresenta il gateway strategico per sbloccare il potenziale analitico della piattaforma, rendendola il primo passo critico in qualsiasi progetto basato sui dati.

Una volta effettuato l'accesso all'interfaccia web di Splunk, ci si trova di fronte alla schermata principale. La barra di navigazione superiore contiene le voci principali di gestione, tra cui "Administrator", "Messaggi" e "Impostazioni". Per il nostro scopo, l'opzione di interesse è il pulsante **"Aggiungi dati"**, chiaramente visibile nel pannello di sinistra.



Sebbene il menu "Impostazioni" offra accesso a configurazioni più avanzate della piattaforma — come la gestione degli utenti, la creazione di indici, la configurazione del clustering e la gestione delle licenze — per l'acquisizione standard di una nuova sorgente dati, la via più diretta è utilizzare l'apposito pulsante.

Una volta cliccato su "Aggiungi dati", si accederà alla schermata successiva, dove dovremo scegliere la modalità di acquisizione più adatta alle nostre esigenze.

2. Selezione della Modalità di Acquisizione Dati

La scelta del metodo corretto per inviare i dati a Splunk è un passaggio strategico, poiché determina come i dati vengono raccolti (se in modo continuativo o come caricamento una tantum) e da dove provengono (dalla macchina locale o da sistemi remoti). La piattaforma presenta tre opzioni principali, ciascuna con un caso d'uso specifico.

- **Carica:** Questa modalità è ideale per l'upload una tantum di file dal proprio computer. È perfetta per analizzare dati storici, file di log archiviati o per testare il parsing di file strutturati come i CSV.
- **Monitora:** È la scelta prediletta per il monitoraggio continuo di sorgenti dati "vive" che risiedono

sulla stessa istanza in cui è installato Splunk. Supporta una varietà di input, tra cui File, porte TCP/UDP, chiamate HTTP, script personalizzati e query WMI.

- **Inoltra:** Questa opzione è progettata per le architetture distribuite. Viene utilizzata per configurare Splunk affinché riceva dati da un "Universal Forwarder", ovvero un agente leggero installato su macchine remote che ha il compito di raccogliere e inoltrare i dati a un'istanza centrale di Splunk.

Per questo manuale, l'obiettivo è raccogliere i log generati localmente dalla macchina su cui è installato Splunk. Pertanto, la procedura che seguiremo sarà quella associata all'opzione **"Monitora"**. Selezionando questa opzione, avvieremo una procedura guidata per la configurazione specifica della sorgente dati.

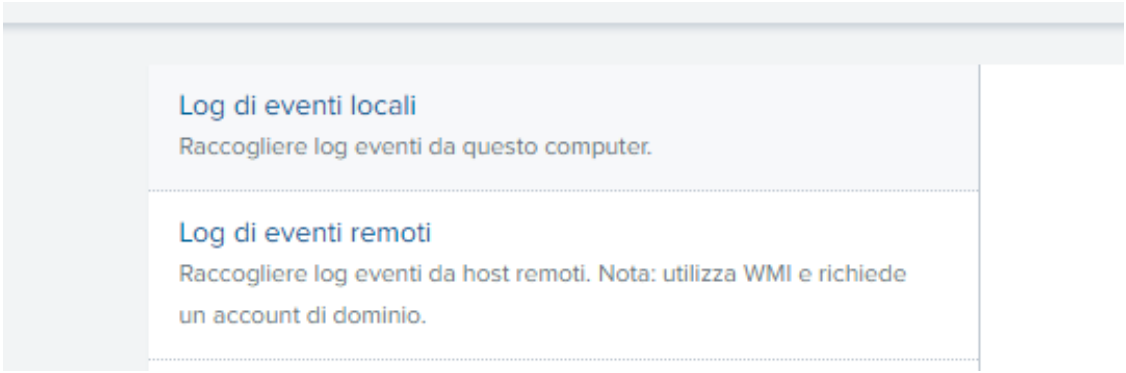
3. Procedura Guidata di Configurazione della Sorgente Dati

Una volta scelta la modalità "Monitora", la piattaforma avvia una procedura guidata per la configurazione dell'input. Questo wizard semplifica un processo che altrimenti potrebbe risultare complesso, assicurando che tutti i parametri fondamentali vengano definiti in modo corretto per garantire un'acquisizione dati efficace.

3.1 Scelta del Tipo di Dati da Monitorare

Il primo passo consiste nel specificare la categoria esatta di dati che si desidera monitorare sulla macchina locale. L'interfaccia propone diverse opzioni, tra cui le più comuni per un sistema Windows.

- **Log di eventi locali:** Questa opzione permette di "Raccogliere log eventi da questo computer."
- **Log di eventi remoti:** Consente di raccogliere log da host remoti, ma come specificato dalla nota, "utilizza WMI e richiede un account di dominio."

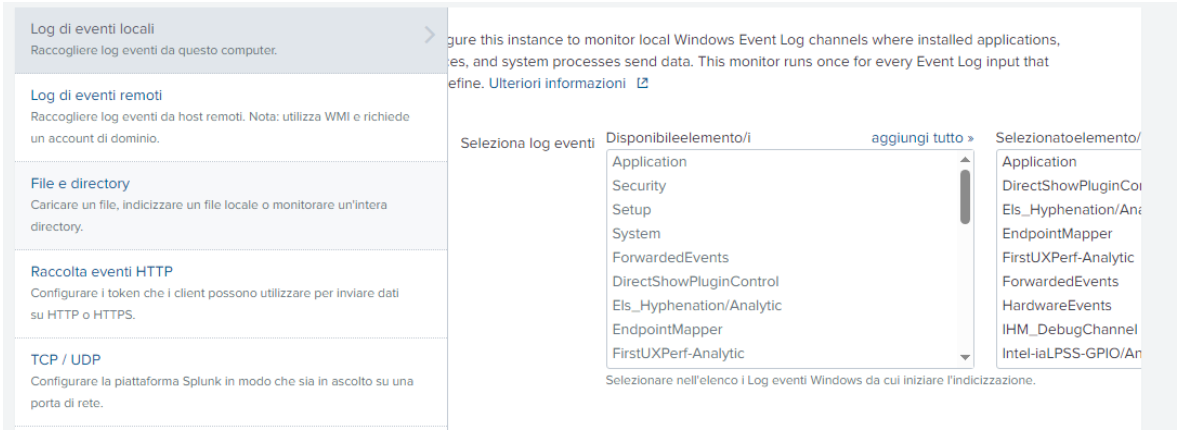


Dato che il nostro obiettivo è monitorare la macchina su cui stiamo operando, selezioneremo **"Log di eventi locali"**.

3.2 Selezione dei Canali di Log

Dopo aver specificato che intendiamo monitorare i log di eventi locali, Splunk ci chiede di definire quali specifici canali di log di Windows vogliamo includere nella raccolta. Questo permette una configurazione granulare, evitando di indicizzare dati non necessari.
L'interfaccia di selezione è suddivisa in due colonne:

- **Disponibile:** Elenca tutti i canali di log rilevati sul sistema. Tra i più comuni troviamo `Application`, `Security`, `Setup`, `System` e `ForwardedEvents`.
- **Selezionato:** Mostra i canali che sono stati scelti per il monitoraggio.



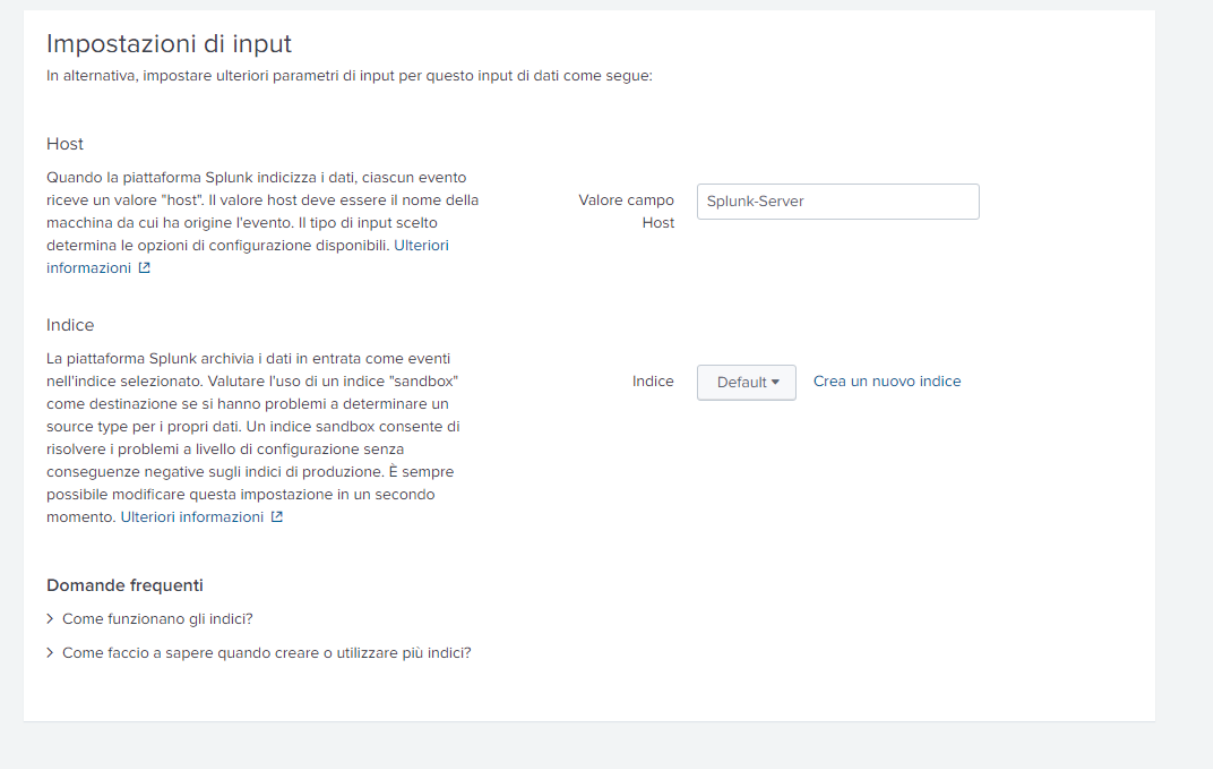
L'utente può selezionare uno o più canali dalla lista "Disponibile" per spostarli in "Selezionato", garantendo così una raccolta dati mirata ed efficiente, focalizzata esclusivamente sulle informazioni di interesse per le proprie analisi. In questo caso, usiamo “aggiungi tutto” per aggiungere tutti i canali proposti.

3.3 Definizione delle Impostazioni di Input

In questa fase vengono definiti alcuni metadati cruciali che Splunk assocerà a ogni evento acquisito.

Questi parametri sono fondamentali per contestualizzare i dati, renderli facilmente ricercabili e organizzarli correttamente all'interno della piattaforma.
I due parametri principali da configurare sono:

- **Host:** Questo valore identifica univocamente la macchina da cui provengono i dati. Come descritto nell'interfaccia, "il valore host deve essere il nome della macchina da cui ha origine l'evento". Nel nostro caso, il valore impostato è Splunk-Server. Definire correttamente l'host è essenziale in ambienti con più server, in quanto permette di distinguere i log di un web server 'WebApp01' da quelli identici provenienti da 'WebApp02'.
- **Indice (Index):** L'indice è il "contenitore" logico, o repository, in cui Splunk archivia fisicamente i dati. La scelta dell'indice è strategica per l'organizzazione dei dati, la gestione dei periodi di conservazione (*retention*) e la definizione dei permessi di accesso per i vari utenti. Per questo esempio, utilizzeremo l'indice predefinito, Default.



L'interfaccia include anche una sezione "Domande frequenti" che fornisce utili approfondimenti sul funzionamento e la gestione degli indici.

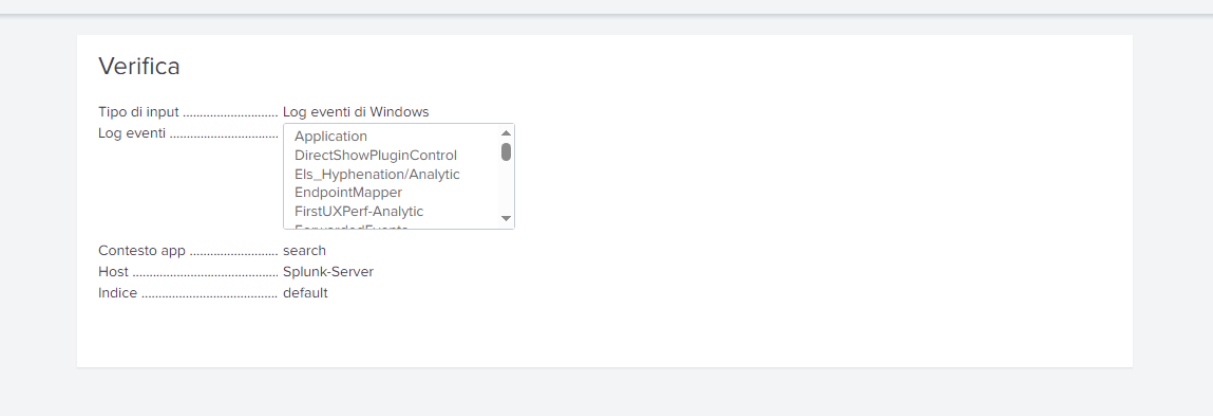
3.4 Riepilogo e Verifica della Configurazione

Prima di finalizzare la creazione dell'input, Splunk presenta una schermata di riepilogo. Questa fase rappresenta un controllo strategico finale che permette di verificare tutte le impostazioni scelte e di correggere eventuali errori prima di attivare la raccolta dati. Questo passaggio è strategico perché previene l'indicizzazione di dati mal configurati, un errore che può essere costoso in termini di utilizzo

della licenza, consumo di storage e tempo necessario per la successiva bonifica.

La schermata di verifica riassume le seguenti configurazioni:

- **Tipo di input:** Log eventi di Windows
- **Log eventi:** Application (e gli altri canali selezionati)
- **Contesto app:** search
- **Host:** Splunk-Server
- **Indice:** default



Avere una visione d'insieme chiara e completa previene configurazioni errate e assicura che i dati vengano acquisiti esattamente come desiderato. Una volta confermata la correttezza di tutti i parametri, si può procedere con la creazione effettiva dell'input.

4. Completamento del Processo e Verifica dei Dati

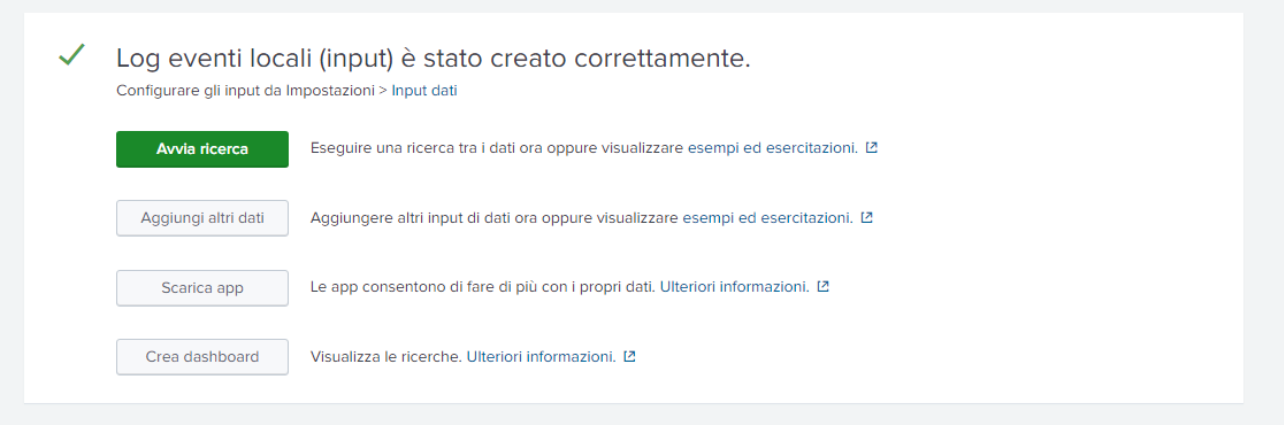
La semplice creazione dell'input non conclude il processo. È fondamentale procedere con una verifica attiva per assicurarsi che i dati stiano effettivamente affluendo nella piattaforma e vengano indicizzati correttamente, pronti per essere analizzati.

4.1 Conferma della Creazione

Al termine della procedura, Splunk visualizza un messaggio di successo che conferma la corretta configurazione dell'input: "**Log eventi locali (input) è stato creato correttamente.**"
A questo punto, l'interfaccia offre un bivio strategico con quattro possibili azioni successive:

- **Avvia ricerca:** L'opzione più immediata per verificare che i dati siano stati acquisiti correttamente, eseguendo una ricerca sugli eventi appena indicizzati.
- **Aggiungi altri dati:** Permette di tornare all'inizio del processo per configurare una nuova sorgente dati.

- **Scarica app:** Consente di esplorare e installare applicazioni dallo Splunkbase per estendere le funzionalità della piattaforma.
- **Crea dashboard:** Avvia il processo di creazione di una dashboard per iniziare a visualizzare i dati raccolti.



Per il nostro scopo di verifica, il passo più logico e immediato è selezionare **"Avvia ricerca"**.

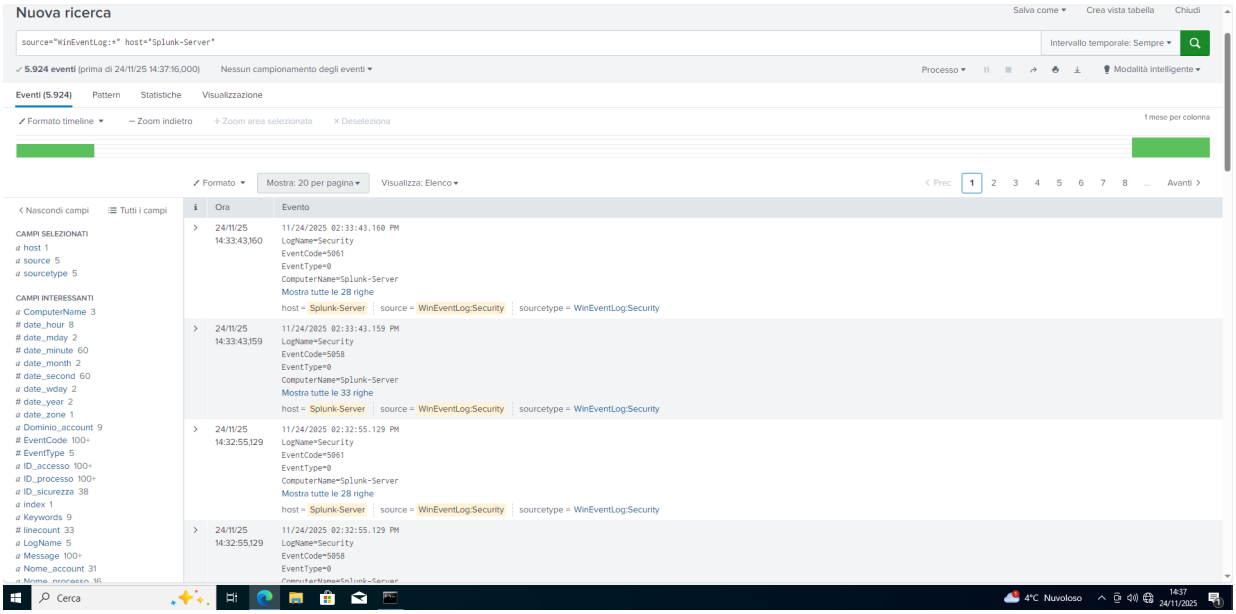
4.2 Ispezione dei Dati tramite Ricerca

Selezionando "Avvia ricerca", si viene reindirizzati all'interfaccia di ricerca di Splunk, che rappresenta la prova finale del successo dell'intera operazione. La piattaforma precompila automaticamente una query per visualizzare i dati appena acquisiti.

La query di ricerca è `source="WinEventLog:*" host="Splunk-Server"`. Analizziamola:

- `source="WinEventLog:*"`: Istruisce Splunk a selezionare tutti gli eventi la cui sorgente corrisponde ai log di eventi di Windows.
- `host="Splunk-Server"`: Filtra ulteriormente i risultati per mostrare solo gli eventi provenienti dall'host che abbiamo specificato durante la configurazione.

Il risultato della ricerca, "**5.924 eventi** trovati", conferma che un volume significativo di dati è già stato indicizzato con successo.



L'interfaccia dei risultati mostra ogni singolo evento con il suo timestamp, l'host e la sorgente. Fondamentalmente, Splunk assegna anche un campo sourcetype (ad es. WinEventLog:Security), che classifica automaticamente la tipologia di dato per facilitare ricerche, analisi e data modeling. La visualizzazione di questi eventi e dei campi estratti non solo conferma che la sorgente dati è attiva e funzionante, ma dimostra anche che Splunk sta correttamente analizzando (*parsing*), indicizzando e rendendo ricercabili i dati come previsto.

Conclusione

In questo manuale abbiamo seguito l'intero flusso di lavoro per l'acquisizione di una nuova sorgente dati in Splunk. Partendo dalla navigazione iniziale nell'interfaccia, abbiamo selezionato la modalità di monitoraggio più adatta, eseguito una configurazione granulare della sorgente specificando canali di log, host e indice, e infine verificato il successo dell'operazione tramite l'interfaccia di ricerca.

La padronanza di questo processo di *data ingestion* è una competenza fondamentale e propedeutica a tutte le funzionalità più avanzate offerte da Splunk. Con una solida base di dati correttamente acquisita, è ora possibile procedere con la creazione di dashboard interattive, l'impostazione di alert proattivi per il monitoraggio in tempo reale e la generazione di report complessi per l'analisi storica.