

I dati di Moodle al servizio della **didattica**

Reportistica scalabile con OpenSearch per docenti, tutor e management nell'esperienza dell'alleanza Across

Stefano Bonomi, Boubakr Injarn

Università degli Studi di Udine

{stefano.bonomi, boubakr.injarn}@uniud.it



Co-funded by
the European Union



Across

Un'alleanza di dieci università in dieci paesi europei, dedicata alla cooperazione transfrontaliera. Il modo in cui si presenta: **dieci università, un campus senza confini.**

L'eCampus

Uno spazio digitale comune per comunicazione, didattica e ricerca, destinato a diventare **il cuore digitale dell'alleanza**. Non sostituisce gli LMS degli atenei: li affianca.



ecampus.
across-alliance.eu

Rendicontare e misurare

Misurare l'uso dei servizi è un'attività prevista dal progetto, e riguarda anche quello che succede dentro Moodle.

Gli strumenti nativi di Moodle nascono per **consultare** i log di un sito, non per analizzarne anni su dieci piattaforme.

Il monitoraggio è **ricorrente**. Ogni estrazione è una richiesta a chi gestisce i sistemi, un'attesa, una verifica, spesso una seconda richiesta. Di qui la domanda: perché non **un punto unico dove chi amministra trova i dati in tempo reale e si fa da sé le proprie analisi?**

Dieci università, un solo eCampus



- Nove atenei hanno **target di indicatore assegnati**; Leopoli partecipa come partner associata e non ne ha.
- L'eCampus ospita i servizi comuni: SSO, catalogo corsi, bacheca tesi, dashboard personale, guida alla mobilità.
- Gli **LMS restano quelli dei singoli atenei**, in larga parte Moodle. Nessuno migra, tutti vanno misurati.
- Anche l'**accesso a un corso dentro l'LMS** conta come accesso a un servizio: è la parte che riguarda direttamente Moodle.

Il problema tecnico nasce da qui: **misurare in modo omogeneo dieci piattaforme che non condividono un database.**

COSA VA MISURATO

Il finanziamento si rendiconta in numeri

- Quante **persone distinte** usano i servizi dell'eCampus in un anno, contando ciascuna una volta sola.
- Quanti **accessi ai servizi**, contando ogni uso effettivo e non ogni clic: un login, l'invio di un modulo, l'apertura di un corso nell'LMS.
- Quanti **servizi sono attivi** sull'eCampus e quanti di questi attingono al knowledge graph.
- Quanto **personale ha seguito** la formazione sul knowledge graph.

centinaia di migliaia

gli accessi ai servizi da dimostrare per ateneo, ogni anno

migliaia

le persone distinte da riconoscere, sempre per ateneo e per anno

milioni

gli elementi nel knowledge graph condiviso dall'alleanza

Ogni numero va prodotto **per ogni ateneo, ogni anno, con una prova verificabile**, e serve durante l'anno, non solo alla scadenza.

Il registro di Moodle non è un archivio di analisi

DI COSA PARLIAMO

Quando qualcuno usa Moodle, le azioni rilevanti diventano **eventi**: un login, l'apertura di un corso, la consegna di un compito. Il log store standard, quello attivo di serie, li registra in un'unica tabella, `logstore_standard_log`. È ciò che alimenta i report sui log, ed è la fonte più ricca sull'uso reale della piattaforma.

IL LIMITE STRUTTURALE

Cresce, e di serie non si svuota

Ogni evento aggiunge una riga. Una scadenza esiste, ma il valore predefinito è **non cancellare mai**: su installazioni medio grandi si arriva a decine o centinaia di milioni di righe. Un database relazionale le conserva bene, ma per contarle deve scorrerle.

L'EFFETTO PRATICO

Il report rallenta la didattica

Le interrogazioni sono sincrone e girano sullo stesso database che serve le lezioni. Non è un timore teorico: il tracker di Moodle documenta **timeout nella consultazione dei log** su installazioni grandi, report di partecipazione lenti già al caricamento dei filtri, ed esportazioni che esauriscono la memoria.

IL LIMITE DI FONDO

Nessuna proattività, nessuna federazione

Tutto è consultazione a posteriori: nessuna notifica quando qualcosa si muove. E dieci atenei restano dieci database che non si parlano.

DUE PAROLE SUL MOTORE

OpenSearch, in breve



FINO AL 2021

Motore di ricerca e analisi distribuito, per anni rilasciato con licenza Apache 2.0.

GENNAIO 2021
CAMBIO DI
LICENZA



APRILE 2021
NASCE IL FORK



DAL 2021

Fork della versione 7.10.2, avviato da AWS per tenere il motore sotto Apache 2.0.

COME FUNZIONA

Indicizza, non scorre

Ogni evento diventa un documento indicizzato campo per campo. Conteggi e raggruppamenti si risolvono sugli indici, senza leggere una riga alla volta.

CHI LO GOVERNA

Sotto la Linux Foundation

Da settembre 2024 il progetto è retto dalla OpenSearch Software Foundation: governance aperta a più aziende, non più di una sola.

PERCHÉ CI INTERESSA

Apache 2.0, davvero

Si usa, si modifica e si ridistribuisce senza restrizioni. Per un progetto finanziato a termine, la licenza è parte della sostenibilità.

STESSA DOMANDA, DUE MOTORI

Quanto pesa davvero la differenza

LA STESSA INTERROGAZIONE SU 100 MILIONI DI RIGHE

Quanti utenti distinti?

`COUNT(DISTINCT utente)`, lo stesso conto delle persone distinte

MYSQL

389
SECONDI

ELASTICSEARCH

2,0
SECONDI

DIVARIO

193x

I dieci valori più frequenti

`GROUP BY ... ORDER BY COUNT DESC LIMIT 10`, la forma di "i dieci corsi più visitati"

464
SECONDI

1,2
SECONDI

379x

Cercare una parola dentro un testo

`WHERE campo LIKE '%parola%'`, il filtro che nessun indice tradizionale può sfruttare

437
SECONDI

13,2
SECONDI

33x

184x

divario mediano sulle 42 interrogazioni confrontabili

40 su 42

le interrogazioni in cui Elasticsearch vince. Non vince sempre

1,7x

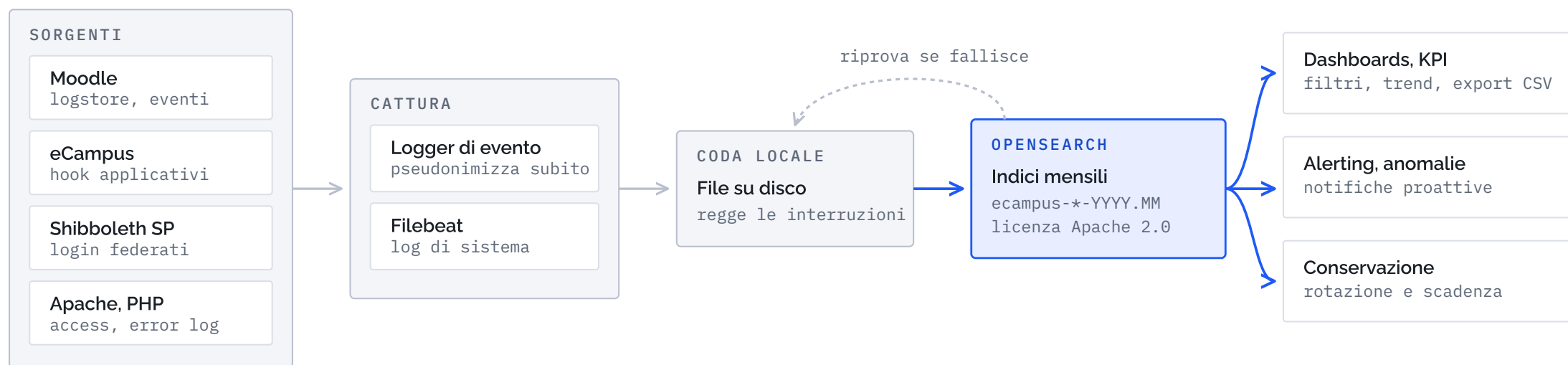
lo spazio su disco in più che Elasticsearch richiede: 157 GB contro 94

Sulla ricerca testuale il divario si restringe, perché un jolly iniziale mette in difficoltà anche un motore di ricerca. Resta comunque la differenza fra attendere tredici secondi e attendere sette minuti. **E il guadagno si paga in spazio.** È il compromesso di fondo: si duplica il dato e si spende disco, per non spendere tempo di calcolo sul database che serve la didattica.

Fonte: ClickBench, benchmark pubblico e riproducibile per database analitici, benchmark.clickhouse.com. Stessa macchina (AWS c6a.4xlarge: 16 vCPU AMD EPYC a 3,6 GHz, 32 GiB di RAM, disco EBS), stesso dataset di 99.997.497 righe reali, configurazioni predefinite, nodo singolo, interrogazioni in sequenza. Rilevazioni MySQL 11/07/2025 ed Elasticsearch 02/09/2025. Tempi: la più rapida di tre esecuzioni. Mediana e conteggio sono calcolati da noi sui risultati grezzi pubblicati. ClickBench non pubblica una voce OpenSearch: il confronto usa Elasticsearch, da cui OpenSearch deriva. Il dataset è di web analytics, non log di Moodle.

LA PROPOSTA

Disaccoppiare la misura dalla produzione



Il database di produzione **non viene mai interrogato per fare reportistica**. Gli eventi lo attraversano una volta sola, in scrittura, e proseguono altrove.

PERCHÉ CI SI PUÒ CONTARE

Un numero che vale come prova

SE IL SISTEMA DI ANALISI SI FERMA

Il sito continua a funzionare

La raccolta è disaccoppiata dalla navigazione: se OpenSearch non risponde, gli eventi restano in coda sul server e nessuno studente se ne accorge. Un guasto al sistema di misura non diventa mai un guasto alla didattica.

SE LA RETE CADE

Nessun dato va perso

Gli eventi aspettano su disco e vengono inviati appena il servizio torna disponibile. Le misure di un'interruzione di mezza giornata si recuperano da sole, senza intervento manuale.

SE UN INVIO VIENE RIPETUTO

Il conteggio non raddoppia

Ogni evento nasce con un identificativo proprio: se dopo un errore di rete viene rispedito, sovrascrive sé stesso invece di contarsi due volte. Quando il numero finisce in una rendicontazione, questa non è una raffinatezza.

COSA CONTA E COSA NO

Le regole sono scritte, non interpretate

Le azioni amministrative come logout, cancellazioni e bozze vengono registrate ma escluse dal conteggio, perché non sono usi deliberati di un servizio. La definizione sta nel codice, uguale per tutti e per tutti gli anni.

COSA VEDE CHI

La rendicontazione diventa un filtro

- Dashboard interattive **filtrabili per corso, categoria, ruolo e periodo**, consultabili in autonomia senza scrivere una query.
- Un pannello dedicato agli indicatori: valore corrente, avanzamento sul target e **confronto con la quota di anno già trascorsa**, con il dettaglio per partner e per servizio, il trend mensile e l'export in CSV.
- Ogni evento porta con sé l'elenco degli indicatori che alimenta. Chi rendiconta **applica un filtro**, non ricostruisce una query.
- Gli stessi dati servono a due pubblici diversi: il docente guarda il suo corso, chi rendiconta guarda l'ateneo.

IL GUADAGNO MISURABILE

Da minuti a millisecondi

Le aggregazioni che sul database relazionale richiedono minuti si risolvono in frazioni di secondo anche su milioni di eventi, e senza toccare la produzione.

EFFETTO COLLATERALE UTILE

La tabella dei log si può alleggerire

Se la storia lunga vive altrove e la conservazione è governata da regole esplicite, la tabella dei log in produzione non deve più essere l'archivio di tutto.

POTENZIALITÀ

Dal report consultato al monitoraggio continuo

NOTIFICHE AUTOMATICHE

Gli avvisi arrivano da soli

Avvisi su soglie di **inattività degli studenti**, picchi anomali di accesso, errori applicativi ricorrenti, tempi di risposta che peggiorano. È il salto vero rispetto a qualunque plugin di reportistica.

RILEVAMENTO ANOMALIE

Soglie che non devi tarare

Modelli nativi che apprendono la stagionalità di un ateneo, dalle sessioni d'esame alle pause didattiche, invece di richiedere una soglia fissa che sarebbe sbagliata undici mesi su dodici.

GESTIONE DEL CICLO DI VITA

La conservazione come regola

Rotazione mensile degli indici e passaggi automatici fino alla cancellazione, dichiarati una volta sola. La conservazione dei dati diventa una regola scritta e verificabile, non uno script che qualcuno ricorda di aver messo in crontab.

LICENZA APACHE 2.0

Nessun vincolo di fornitore

Scala aggiungendo nodi, e non c'è una licenza da rinnovare. In un progetto finanziato a termine non è un dettaglio economico: è la condizione perché il risultato sopravviva al progetto.

Nessuna informazione personale nei dati di evento. Se un dato non serve a un indicatore, non viene raccolto.

Utente, indirizzo IP e sessione diventano impronte irreversibili. Nulla in chiaro lascia l'applicazione.

Gli indici hanno una scadenza dichiarata. La conservazione è una regola scritta, non una consuetudine.

Trasporto cifrato, credenziali ristrette,
pannello riservato agli amministratori.

Chi entra sa fin dal primo accesso che l'uso dei servizi viene conteggiato, e perché.

Con dieci atenei le responsabilità vanno concordate per iscritto. È lavoro in corso.

across
EUROPEAN UNIVERSITY
FOR CROSS-BORDER
KNOWLEDGE SHARING

moodle
moot
ITALIA 2026

TRE COSE DA PORTARE A CASA

Grazie

1 I dati diventano consultabili da chi non scrive query

Chi deve guardare entra e trova, in tempo reale: nessuna richiesta, nessuna attesa. E le notifiche arrivano senza doverle chiedere.

2 La privacy si progetta all'origine

Sostituire l'identità quando l'evento nasce costa meno che ripulire a valle, e copre più principi del solo nascondere il nome.

3 Lo stack è aperto

Nessuna licenza da rinnovare quando il progetto finisce.

Stefano Bonomi, Boubakr Injarn

Università degli Studi di Udine

{stefano.bonomi, boubakr.injarn}@uniud.it



Co-funded by
the European Union

across
EUROPEAN UNIVERSITY
FOR CROSS-BORDER
KNOWLEDGE SHARING

aium
associazione italiana utenti moodle.org