



UNIVERSITÀ DEGLI STUDI DI MILANO
DIPARTIMENTO DI INFORMATICA

Corso di Laurea Magistrale in
Scienze e Tecnologie dell'Informazione

Mole.io

un Sistema per la Gestione Centralizzata
dei Log Applicativi

RELATORE

Prof. Ernesto Damiani

TESI DI LAUREA DI

Federico Gandellini

CORRELATORE

Matr. 703156

Dott. Emanuele DelBono

Anno Accademico 2014/2015

Ringraziamenti

Un grazie a ...

Indice

Introduzione	3
1 Log:	
Contesti e Problematiche	6
1.1 Trattare gli Errori Applicativi	8
1.2 La Centralizzazione	10
1.3 Business Intelligence	12
2 Log: Software e Applicazioni	14
2.1 Prodotti e Soluzioni sul Mercato	15
2.2 Una Nuova Applicazione: Mole.io	27
3 Metodologie di Sviluppo	29
3.1 User Story	33
3.2 Il Test Driven Development	36
4 Tecnologie Utilizzate	39
4.1 Node.js	40
4.1.1 Le operazioni asincrone	42
4.1.2 Le callback	43
4.1.3 La Storia	46
4.1.4 Npm e Moduli	47

INDICE	2
4.2 RabbitMQ	52
4.3 MongoDB	56
4.3.1 Fronteggiare le Richieste	59
4.4 AngularJS e Altre Tecnologie di Frontend	62
4.4.1 Gestione delle Dipendenze	63
4.5 Strumenti per il Deploy	64
5 Mole.io	65
5.1 Architettura del Sistema	67
5.1.1 CQRS ed Estensibilità	68
5.1.2 Mole	69
5.1.3 Mole Suit	71
5.2 Autenticazione degli Utenti	73
5.3 Scalabilità e Affidabilità	74
5.4 Problematiche di Sviluppo	75
6 Configurazioni e Benchmark	76
Conclusioni e sviluppi futuri	77
Bibliografia	79

Introduzione

In questa tesi si descriverà Mole.io: un sistema centralizzato per la raccolta e l'aggregazione di messaggi provenienti da applicazioni remote.

Durante il loro ciclo di lavoro o *processing*, le applicazioni software eseguono operazioni significative o entrano in situazioni di errore. In questi casi è importante che le persone che hanno in carico la gestione di questi sistemi, siano informate dell'accaduto in modo da operare scelte opportune o applicare le dovute correzioni (*bugfix*).

Gli sviluppatori sono soliti utilizzare messaggi di tracciamento (*log*) per stampare a video o salvare in *file* stati significativi delle applicazioni. I messaggi più frequenti riportati nei log sono quelli relativi a situazioni di errore (*Exception* e *Stack Trace*).

L'approccio comune alla creazione e gestione dei log presenta la criticità specifica della *località*, poiché tipicamente questi file vengono salvati nella stessa macchina sulla quale sta operando l'applicazione.

All'aumentare del numero di applicazioni da gestire e del numero di macchine in produzione, capita spesso che i server siano in luoghi geograficamente distanti tra loro. Questa situazione rende evidente la difficoltà di ottenere un *feedback* veloce dello stato di ogni software e delle eventuali situazioni di errore in cui le applicazioni si trovano.

Mole.io cerca di risolvere il problema facendo in modo che i software che

lo utilizzano, siano in grado di inviare le informazioni che ritengono significative ad un server centrale, il quale le raccoglie, le cataloga e le aggrega per essere facilmente supervisionate da parte degli sviluppatori.

L'esigenza di una applicazione per la centralizzazione dei log nasce da CodicePlastico, una azienda con sede a Brescia, che si occupa di realizzare applicazioni su misura per i propri clienti.

La gestione di un gran numero installazioni dislocate sul territorio e di molte realtà aziendali con esigenze differenti ha reso, per CodicePlastico, particolarmente complesso il tracciamento dello stato di ogni software in produzione. Questa situazione ha spinto l'azienda a decidere di dotarsi di un sistema centralizzato in grado di collezionare i log prodotti dai diversi applicativi, organizzarli e catalogarli in modo automatico.

Il nuovo approccio permette agli sviluppatori di identificare, in breve tempo, il manifestarsi di un malfunzionamento in qualunque applicazione installata presso uno dei proprio clienti e reagire rapidamente proponendo una azione risolutiva.

CodicePlastico opera nel settore IT avvalendosi di strumenti software di vario genere, sia proprietari, sia *open source*. Per il *deploy* in produzione di Mole.io, si è scelto di utilizzare *Microsoft Azure*, un sistema *PaaS* distribuito con il quale è possibile creare architetture facilmente scalabili per supportare la variabilità del carico di lavoro richiesto al sistema.

Durante la fase iniziale Mole.io sarà utilizzato da un ristretto numero di clienti e applicato solo ad alcune installazioni. Con il passare del tempo, auspicabilmente, questo numero crescerà ed una architettura scalabile permetterà di affrontare le esigenze di reattività e stabilità del sistema.

Nel primo capitolo saranno trattati approfonditamente la tematica dei log, i contesti nei quali essi vengono utilizzati e le problematiche legate

alla gestione di questo tipo di soluzione di tracciamento. Vedremo anche come utilizzare i log per ottenere informazioni di supporto alla *business intelligence*.

Il secondo capitolo riporterà un elenco dei principali *software* per la gestione centralizzata dei log presenti sul mercato e delle soluzioni *Open Source* che sono state prese a modello per la realizzazione di Mole.io. Descriveremo ogni applicazione e mostreremo come Mole.io possa essere una soluzione innovativa sotto svariati punti di vista.

I due capitoli seguenti permetteranno di approfondire i dettagli tecnici delle metodologie di sviluppo applicate durante il *design* del software e alcune tra le principali tecnologie utilizzate per la realizzazione del sistema.

Il quinto capitolo descriverà la struttura di Mole.io e le varie componenti software che rendono l'applicazione scalabile e garantiscono l'alta accessibilità della soluzione.

Nel sesto capitolo verrà mostrato in modo oggettivo, con *benchmark* e *stress test* il comportamento di Mole.io all'aumentare del carico di lavoro e sarà dimostrato come le soluzioni di design applicate garantiscano buone *performance*, anche in condizioni critiche di traffico.

Infine verranno discussi i risultati ottenuti e saranno proposte alcune interessanti funzionalità che trasformeranno Mole.io dall'attuale *proof of concept* ad un vero e proprio servizio.

Capitolo 1

Log:

Contesti e Problematiche

Prima di entrare nello specifico delle tematiche trattate, è necessario riprendere e chiarire alcuni concetti chiave che verranno utilizzati ripetutamente nel seguito della tesi.

Si definisce *processo* un programma in esecuzione e *log* l'insieme dei messaggi prodotti, a fini informativi, da tale processo.

I messaggi di log posso essere di vario tipo. È usanza comune caratterizzare ogni messaggio con un livello di gravità (*severity*) permettendo così una rapida identificazione degli errori critici allo scopo di rendere repentino l'intervento di riparazione dell'applicazione.

Benché esista un protocollo riconosciuto a livello internazionale e adottato negli ambienti *Unix-like* chiamato *SysLog*, nell'ambito dello sviluppo di applicativi, non esistono norme vincolanti per la strutturazione dei log stessi. Questo accade sia perché SysLog non rappresenta uno standard rigidamente definito, sia perché l'organizzazione delle informazioni contenute nei log, è spesso delegata agli sviluppatori, i quali implementano questa funzionalità

nel modo più conveniente rispetto alle specifiche esigenze dell'applicazione in costruzione.

All'interno dei log, di conseguenza, troveremo informazioni diversificate in base al caso d'uso, ma tra le più frequenti possiamo citare:

- dati specifici del sistema nel quale l'applicazione è in esecuzione;
- dati relativi all'utente che sta utilizzando l'applicazione;
- dati relativi allo stato del sistema in un preciso istante temporale;
- un marcatore temporale (*timestamp*)
- un messaggio in linguaggio naturale, significativo, che lo rende immediatamente identificabile tra altri;

1.1 Trattare gli Errori Applicativi

Il salvataggio dei log, come abbiamo anticipato, è una operazione molto comune nei software, ma diventa fondamentale quando si vuole monitorare lo stato interno di una applicazione in esecuzione, con l'obiettivo di essere informati riguardo alle situazioni di errore nelle quali quest'ultima incorre.

Salvare le informazioni relative alle situazioni di errore è importante per gli sviluppatori. Questa operazione permette, infatti, di velocizzare l'individuazione di errori (*bug*) nel flusso di lavoro del programma e, di conseguenza la loro risoluzione (*bugfix*).

Il salvataggio dei log avviene tipicamente su uno o più file di testo presenti nella stessa macchina nella quale sta funzionando l'applicazione. A seconda del tempo di esecuzione di una applicazione e della frequenza con la quale essa produce messaggi di log, questi file possono diventare molto grandi.

File di considerevoli dimensioni sono altamente complessi da gestire da parte degli addetti ai lavori. La problematica più evidente diviene infatti trovare informazioni significative all'interno di questa grande mole di dati. Questo processo richiede infatti tempo e attenzione in situazioni di emergenza, nelle quali il ripristino del sistema deve avvenire nel modo più rapido possibile.

Il problema dei file di grandi dimensioni non riguarda esclusivamente la scansione sequenziale delle informazioni, risulta infatti difficoltosa anche l'individuazione del messaggio prodotto a fronte di una criticità e la correlazione di quest'ultima allo stato del sistema nell'istante in cui è stata generata.

L'individuazione degli errori non riguarda esclusivamente l'analisi dei log nell'istante in cui il malfunzionamento si è manifestato, bensì richiede

di comprendere la catena di eventi pregressi, non sempre palese, che ha portato il sistema nella condizione di errore. La capacità dell'analista sta nel riconoscere pattern ricorrenti che generano l'intera situazione.

Un'ulteriore complicazione dovuta alla dimensione eccessiva dei file di log non riguarda solo il riconoscimento delle cause di un problema ma anche l'individuazione di criticità simili occorse in istanti temporali differenti. Questo processo, noto come *clustering*, diviene ovviamente oneroso in termini di tempo, al crescere delle dimensioni del log.

L'analista ha anche un'altra incombenza: verificare che le dimensioni dei file di log non eccedano al punto di compromettere il funzionamento dell'applicazione stessa a causa dell'assenza di spazio su disco fisso.

A livello aziendale il tempo che intercorre tra la scoperta di un bug e il relativo bugfix dovrebbe essere il più possibile contenuto. Spesso purtroppo le eccessive dimensioni dei file di log rendono questa procedura molto costosa.

1.2 La Centralizzazione

L'azienda che ha ospitato lo *stage* di Tesi, Codice Plastico, è una realtà che opera nel mercato IT sviluppando applicazioni su misura di tipo *mobile*, *web* e *desktop*. Il contesto dal quale è nato lo sviluppo di questo progetto è stato la necessità di trovare una soluzione per centralizzare i log in un unico sistema, facilmente accessibile e con un'interfaccia utente *user-friendly*.

Ogni applicazione realizzata e posata dall'azienda, infatti, presenta criticità differenti rispetto alla gestione degli errori:

- Nelle applicazioni web, di norma, i log risiedono su server di proprietà dei clienti, spesso dislocati in aree geografiche differenti.
- Nel caso delle applicazioni mobile, i log risiedono sui device stessi, così come nel caso delle applicazioni desktop.

Tra i numerosi vantaggi della centralizzazione il principale è la riduzione del carico di lavoro dell'analista, il quale può avere accesso ai log senza l'onere di doverli attivamente cercare sui sistemi dei clienti.

Inoltre l'aggregazione dei dati permette all'analista di aver accesso ad informazioni già pre-elaborate, come ad esempio il *clustering* di errori simili avvenuti in istanti temporali differenti con l'evidente semplificazione del processo di riconoscimento delle correlazioni causa-effetto.

Il processo di centralizzazione richiede l'inversione del paradigma di segnalazione degli errori. Si passa da una modalità nella quale è l'analista a dover cercare attivamente i file sui sistemi, ad una in cui sono i sistemi stessi ad inviare i propri log. A questa logica è facilmente integrabile un sistema di notifiche in tempo reale, con lo scopo di rendere repentina la segnalazione dell'errore e il conseguente intervento di ripristino.

In commercio esistono svariati sistemi e *tool* che consentono la gestione delle problematiche legate ai log, come ad esempio l'archiviazione, l'analisi, il *parsing*, il monitoraggio e le segnalazioni. Anche in questo caso la centralizzazione degli strumenti di controllo, evita l'installazione e conseguente manutenzione di numerosi tool su ciascun sistema in produzione.

L'effetto evidente della centralizzazione, della semplificazione dell'accesso ai dati e della loro analisi è la riduzione di costi di manutenzione del software, oltre all'incremento della qualità del prodotto offerto.

1.3 Business Intelligence

La Business Intelligence (BI) indica l'operazione di inferenza di informazioni da una mole di dati, provenienti da fonti differenti nei processi aziendali.

In senso ampio, le fonti di informazioni utili possono essere tra le più disparate, da statistiche di utilizzo dei sistemi informativi, ai dati generati dal funzionamento di software di automazione, al campionamento di flussi di navigazione e modalità di utilizzo dei sistemi.

L'obiettivo della BI è di trarre informazioni e conclusioni utili ai fini aziendali, come l'individuazione delle cause di problemi, la misurazione delle performance, la progettazione di *feature* che potrebbero incrementare la qualità del prodotto o fare previsioni e stime di scenari futuri, sulla base della storia pregressa.

Poiché i log sono strettamente legati al funzionamento delle applicazioni stesse, possono essere utilizzati, oltre che come sistema di controllo dei malfunzionamenti anche come veicolo di raccolta di informazioni utili alla BI.

La BI, infatti, si avvale dei log con diversi fini:

- come strumento per la comprensione e il tracciamento del comportamento degli utenti che operano nel sistema
- per ottenere statistiche di utilizzo del sistema, come ad esempio le funzionalità più utilizzate di un'applicazione o quali aree più visitate di un sito internet.

Poiché i dati utili da collezionare variano in maniera significativa da applicazione ad applicazione, da contesto a contesto, l'idea di costruire un sistema di centralizzazione dei log, che non ponga vincoli nella formulazione

degli stessi, va nell'ottica di poter offrire uno strumento utile di raccolta di informazioni diversificate a supporto alle decisioni.

In altre parole, il sistema può diventare un collettore di informazioni relative al funzionamento del sistema ma non necessariamente legate ai concetti di errore e criticità, divenendo di fatto uno strumento per la gestione strategica dei processi aziendali. Un esempio di tool di questo genere è SpagoBI [1].

Capitolo 2

Log: Software e Applicazioni

Durante la prima fase del progetto si sono analizzate le soluzioni già proposte dal mercato IT, verificando quali tra i prodotti esistenti presentassero la peculiarità della centralizzazione dei log.

In questo capitolo saranno analizzate alcune tra le principali applicazioni dedicate alla raccolta e analisi dei dati. Per ciascuna soluzione si andranno ad indicare le *feature* di ispirazione al progetto, così come i problemi e gli svantaggi identificati.

Lo stato dell'arte della tecnologia in questo settore è attualmente molto ampio: esistono soluzioni complesse e complete che in parte copriranno, con le proprie funzionalità, le necessità individuate in fase di analisi preliminare. Tuttavia si andranno ad analizzare le motivazioni che hanno portato l'azienda ad optare per una soluzione *custom* sviluppata internamente.

2.1 Prodotti e Soluzioni sul Mercato

I prodotti per la gestione centralizzata dei *log* offerti dal mercato sono svariati. Tuttavia, molte di queste soluzioni nascono con l'obiettivo di essere verticalizzate su uno specifico ambito di utilizzo applicativo. Ad esempio alcune soluzioni vengono costruite appositamente per affrontare le problematiche dell'ambito mobile, come altre sono realizzate specificamente per un determinato framework o ambiente di sviluppo.

Per questo motivo sono state escluse dall'analisi soluzioni altamente legate a precisi casi d'uso. Si sono andate, invece, ad individuare le cinque software che affrontano la problematica dei log centralizzati in maniera generica, integrabili con diversi ambienti di sviluppo, e utilizzabili, contemporaneamente, da differenti tipologie di applicazioni.

Airbrake

Airbrake [2] è una delle più conosciute applicazioni per il monitoraggio dei log prodotti da applicazioni mobile. Benché sia diffusa tra le applicazioni per smartphone e dispositivi mobili, offre moduli di integrazione per i principali linguaggi di programmazione e può essere utilizzata anche in ambito web o desktop.

A partire dalla versione 2.0 propone un pannello di gestione, ricerca e aggregazione dei messaggi di errore completamente rinnovato. L'interfaccia grafica è gradevole e ben organizzata.

La *dashboard* principale dell'applicazione riporta tutti i principali errori ottenuti dai software monitorati, questo aiuta il *team* di sviluppo a concentrarsi sulle problematiche più urgenti, massimizzando il valore applicativo percepito dagli utenti e contribuendo ad aumentare la produttività del team stesso.

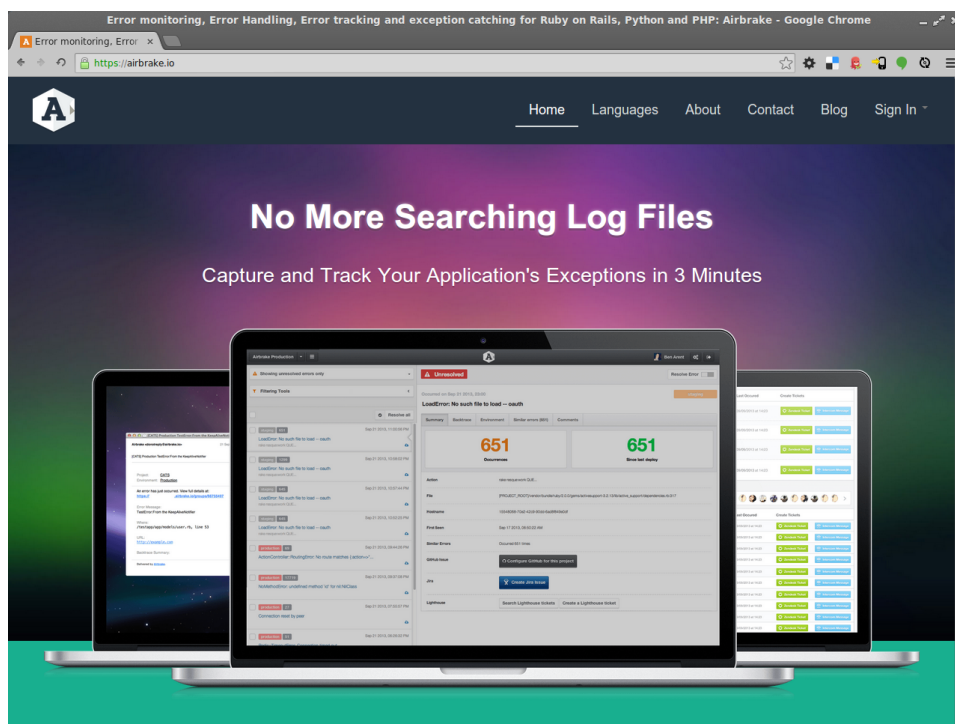


Figura 2.1: Il sito web di Airbrake

Oltre all'aggregazione automatica dei messaggi d'errore, Airbrake possiede moduli di integrazione con i maggiori sistemi di *bug-tracking*. Al momento della ricezione di un messaggio di errore, viene creato *task* a carico di uno sviluppatore che si occuperà poi della gestione della problematica e bugfix.

Airbrake si avvale dell'utilizzo di connessioni sicure SSL al fine di garantire la riservatezza dei dati raccolti.

Il formato dei messaggi di errore interpretato da questo *tool* ha una struttura fissa e non consente l'aggiunta di dati aggiuntivi custom. Questo aspetto rende Airbrake uno strumento poco adatto al tracciamento di messaggi con formati diversi dagli eventi di errore.

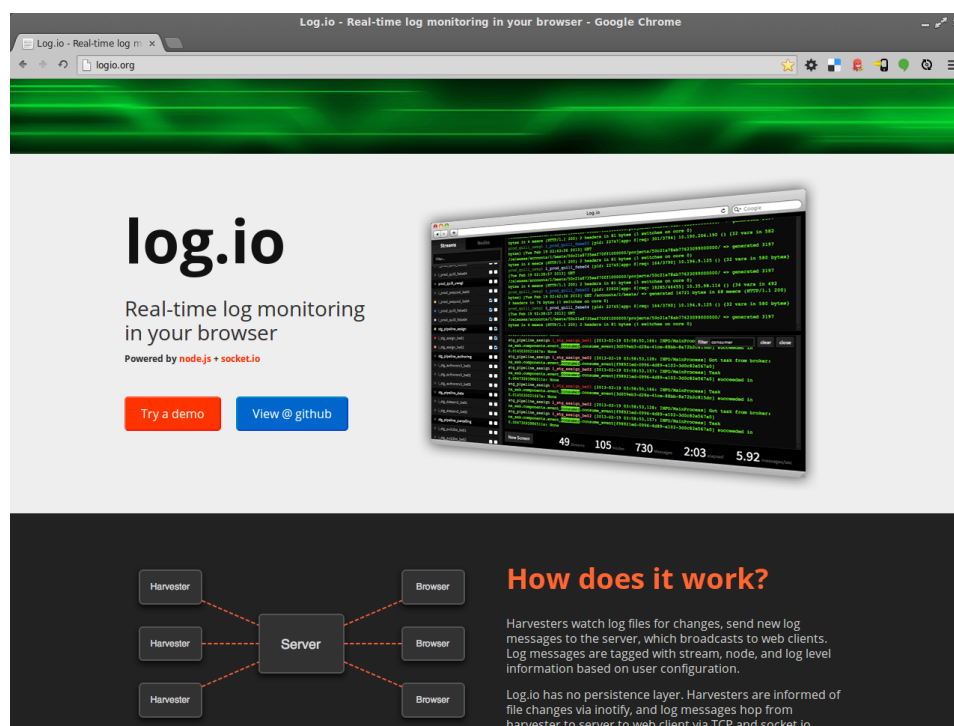


Figura 2.2: Il sito web di Log.io

Log.io

La peculiarità di questo servizio [3] è l'aspetto *realtime*, infatti esso permette di ottenere in tempo reale i log in arrivo dalle applicazioni monitorate.

Log.io non possiede un sistema di persistenza dei dati, i quali vengono salvati dai diversi sistemi monitorati in modo indipendente. Esso si occupa esclusivamente di fornire un punto di raccolta dei messaggi in arrivo dalle diverse sorgenti e di applicare chiavi di filtraggio in tempo reale.

I log vengono gestiti come flussi di dati. L'analista ha la possibilità di utilizzare dei filtri sui dati con lo scopo di limitare la quantità di informazioni riportate, mostrando i soli dati utili all'analisi. Purtroppo questo sistema non esegue alcun tipo di aggregazione dei dati in arrivo e non permette quindi di avere una visione d'insieme della situazione dell'applicazione

monitorata.

L'invio dei messaggi al server, da parte delle applicazioni, avviene tramite messaggi TCP con una formattazione fissa. Questo aspetto contribuisce a rendere Log.io uno strumento difficilmente utilizzabile quando si vogliono notificare informazioni più strutturate di una semplice stringa di testo.

Rollbar

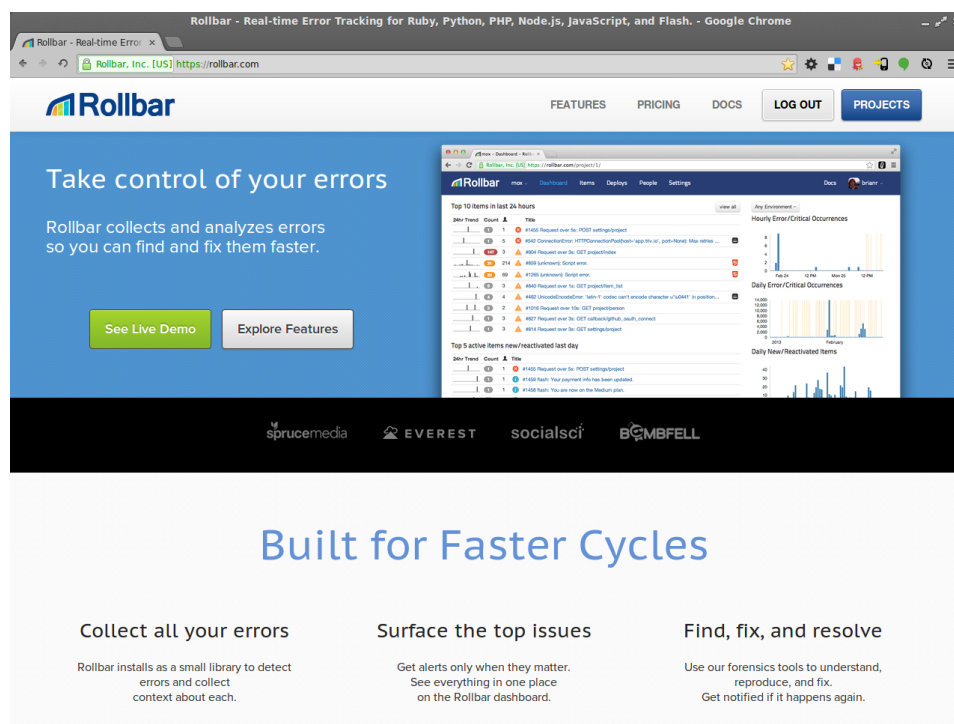


Figura 2.3: Il sito web di Rollbar

Tra le applicazioni a pagamento censite, questa sembra essere la più completa [4], permette infatti di registrare sia messaggi di errore sia messaggi generici.

E' possibile arricchire i messaggi inviati a Rollbar con alcuni dati semplici o strutturati, definibili dall'utente. Le funzionalità di aggregazione dei dati

sui tali messaggi, sono disponibili solo ad alcuni campi obbligatori che il sistema è in grado di gestire. Ad esempio è possibile aggregare i dati per livello di gravità, o per similitudine tra i messaggi, ma non è in grado di fare assunzioni automatiche sui campi custom aggiunti.

Ogni qual volta un sistema viene pubblicato in produzione, è possibile inviare una notifica a Rollbar. Questo permette agli analisti di correlare gli errori in arrivo con la specifica versione dell'applicazione dalla quale essi sono stati generati.

Rollbar offre, inoltre, la possibilità di aggiungere collaboratori ad un progetto, dando la possibilità ad un intero team di sviluppo di condividere le informazioni relative allo stato delle applicazioni monitorate.

In questa applicazione è possibile configurare la notifica via email per alcuni tipi di messaggi e, infine, possiede una modalità di visualizzazione dei dati *realtime*.

Permette infine di notificare gli eventi aggiungendo automaticamente alcuni *task* nei principali sistemi di *time tracking* e *project management* esistenti sul mercato.

Papertrail

E' un tool molto orientato agli amministratori di rete [5]. Nei sistemi *unix-like*, il comando `tail`, permette di stampare in *console* in tempo reale le ultime righe aggiunte ad un file utilizzato da un processo. Analogamente, lo stesso comando in Papertrail, permette di ottenere la coda degli ultimi messaggi ricevuti in tempo reale con una operazione chiamata *Tail and Search*. Papertrail salva periodicamente i log ricevuti e permette di scaricarli, rendendoli così disponibili per eventuali elaborazioni successive.

A livello di analisi automatizzata, questo tool, esegue una classificazione

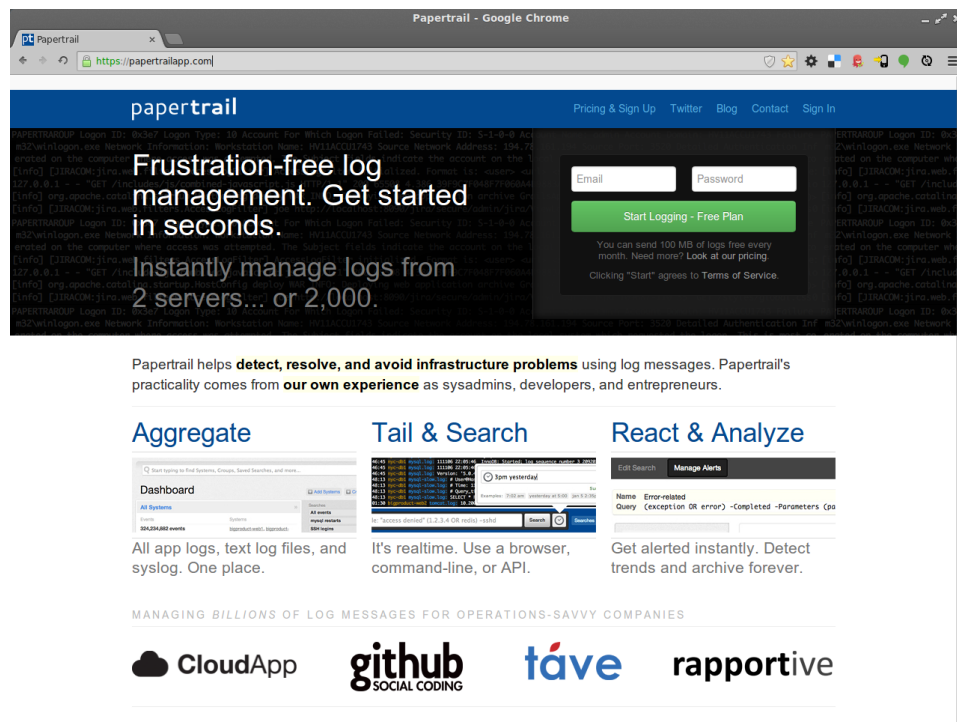


Figura 2.4: Il sito web di Papertrail

dei messaggi ricevuti estrapolando autonomamente la chiave di aggregazione dal formato stesso del messaggio.

Aggrega, inoltre, messaggi di tipo uniforme, come ad esempio messaggi di riavvio del sistema, eventi di *login* degli utenti, o errori di Apache.

E' possibile utilizzare diversi criteri per il filtraggio dei messaggi di errore. Tra questi è di particolare interesse il filtro temporale, è infatti possibile selezionare un intervallo di tempo e ottenere l'elenco dei messaggi ricevuti in quel momento.

L'interfaccia di Papertrail facilita il lavoro dell'analista formattando con colori differenti l'elenco dei messaggi ricevuti. I colori sono correlati al testo in maniera semantica (date, indirizzi ip, ecc...).

Dal punto di vista della personalizzazione dei dati, questa applicazio-

ne, non consente di aggiungere informazioni strutturate ai messaggi inviati. Questo perché il sistema di aggregazione si basa sull'analisi del formato delle stringhe ricevute. Conseguentemente, non solo Papertrail accetta come messaggi esclusivamente i formati stringa e non dati strutturati come accade nelle altre applicazioni prese in esame, ma non è possibile aggiungere ulteriori informazioni ai messaggi stessi.

Fluentd, ElasticSearch e Kibana

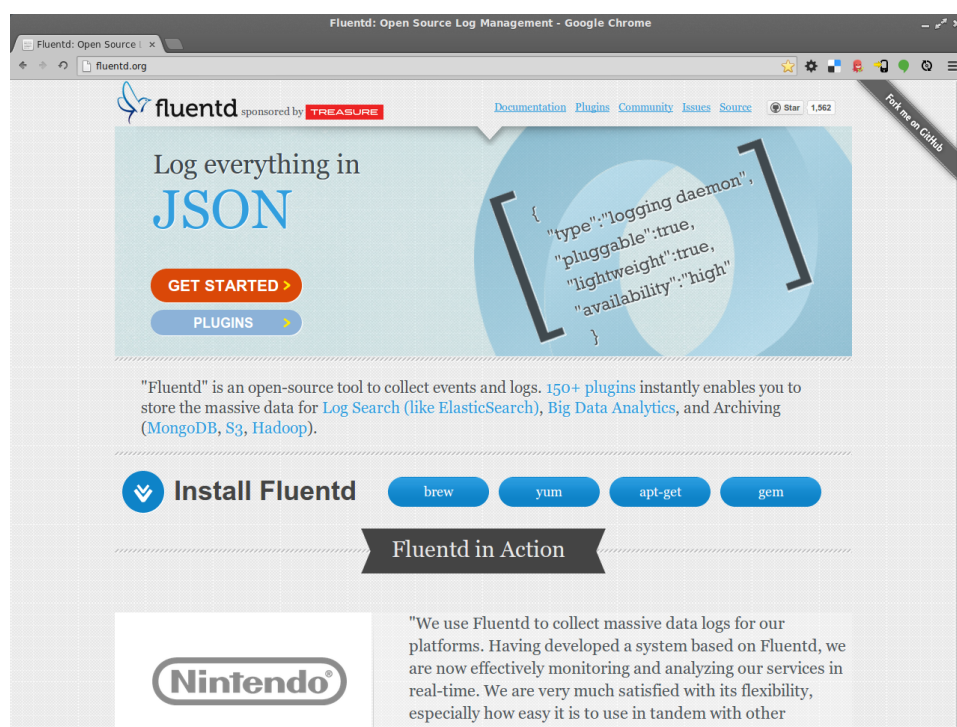


Figura 2.5: Il sito web di Fluentd

Nella analisi comparativa in corso è necessario inserire anche questa architettura composta da tre applicazioni che collaborano sinergicamente al fine di creare uno strumento flessibile e completo per la raccolta, l'archiviazione, la ricerca e l'analisi dei log.

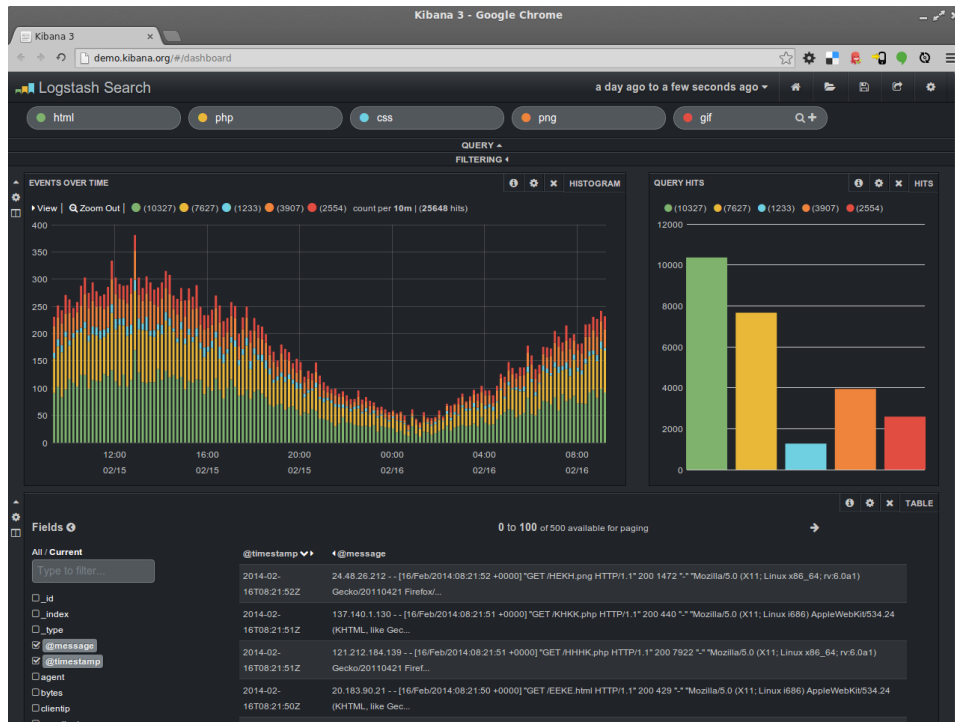


Figura 2.6: Un esempio di interfaccia generata con Kibana

Nonostante questa soluzione software sia la più completa architettura *open source* centralizzata per la raccolta di log, non è stata rilevata da una prima fase di ricerca. Questo si suppone sia dovuto al fatto che ciascun applicativo che compone l'architettura è nato con lo scopo di trattare un sotto-problema specifico, rispetto al tema della gestione centralizzata dei log.

Fluentd E' un collettore di messaggi in formato JSON [6]. Sistemi differenti possono inviare dati all'applicazione che si occupa di redirigerli su *output* differenti, come ad esempio server per l'invio di *email*, file di testo, basi di dati o altri servizi. Grazie ad una vasta gamma di plugin, sviluppati dalla *community*, Fluentd può ricevere i più svariati

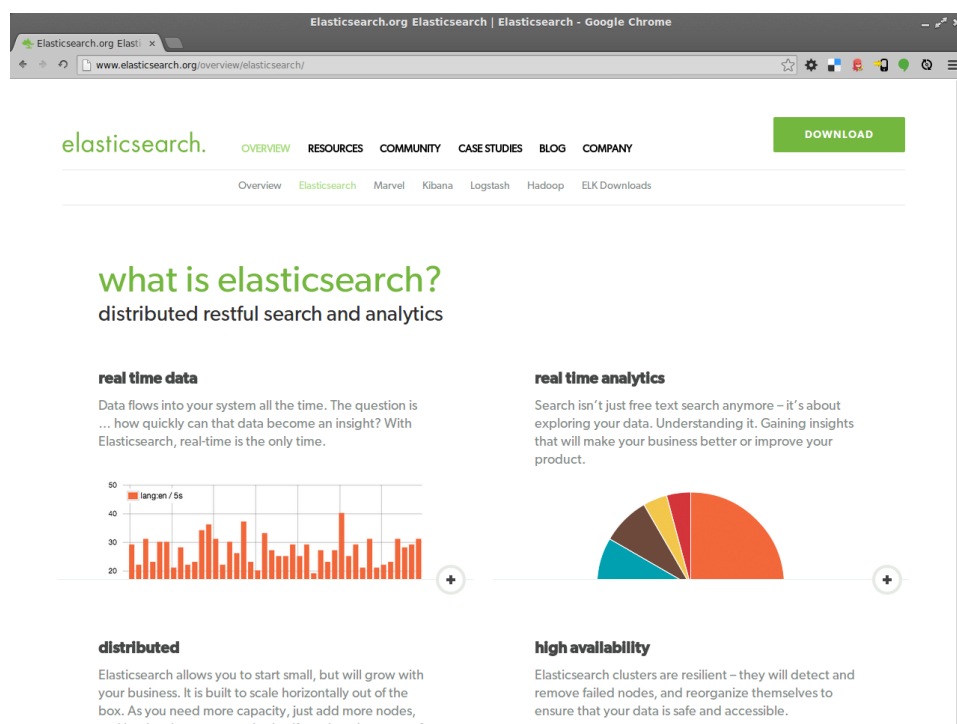


Figura 2.7: Il sito web di Elastic Search

tipi di messaggi che vengono tradotti in JSON dai plugin stessi e in seguito gestiti dall'applicazione. L'intero software è stato ideato per l'installazione su una architettura distribuita, fornendo la possibilità di eseguire un *load balancing* automatico del carico di lavoro dovuto alla ricezione dei messaggi.

Elastic Search E' un sistema distribuito per effettuare ricerche e analisi su grandi quantità di dati [7]. Tra le caratteristiche fondamentali di questo software possiamo annoverare l'alta affidabilità, garantita da un sistema di nodi ridondati, e la possibilità di lavorare in *real-time*. Espone in maniera automatica una interfaccia software che rende molto agevoli le operazioni di filtraggio e aggregazione dei risultati ottenuti. Ad esempio ipotizzando di voler gestire informazioni di log con

la seguente struttura:

```
{  
  timestamp: '2014-01-16T20:19:28.871Z',  
  user: 'mario rossi',  
  message: 'login error'  
}
```

Elastic Search genererà in automatico alcune chiavi di ricerca che permetteranno di filtrare i messaggi ricevuti per `timestamp`, `user` e `message`. Le interfacce per la ricerca e aggregazione dei dati sono rese facilmente fruibili con l'aiuto di *API REST* anch'esse generate automaticamente a partire dall'analisi della struttura dei dati raccolti. La flessibilità fornita da Elastic Search è ottenibile anche grazie all'aiuto di un database documentale *schema less*, che il sistema utilizza per il salvataggio dei dati applicativi ricevuti.

Kibana Si occupa di mostrare a video i dati ottenuti dall'interrogazione dei Elastic Search. Con questo tool [7] è possibile realizzare, in pochi passaggi, svariate tipologie di grafici e tabelle filtrabili e ordinabili.

L'utilizzo di questi tre tool in maniera sinergica permette di costruire un sistema di gestione e analisi dei log efficiente basato sulla architettura mostrata in figura 2.8. I messaggi, provenienti da sorgenti differenti, vengono raccolti dai plugin di Fluentd e tradotti in formato JSON. A questo punto un ulteriore plugin si occuperà di inviare i dati ad Elastic Search il quale li salverà all'interno del proprio database documentale. Sempre Elastic Search avrà il compito di interrogare la base di dati ed esporre tali informazioni aggregate a Kibana attraverso una *REST API*. Questo quindi permetterà

all'analista di creare *report* personalizzabili secondo le esigenze specifiche.

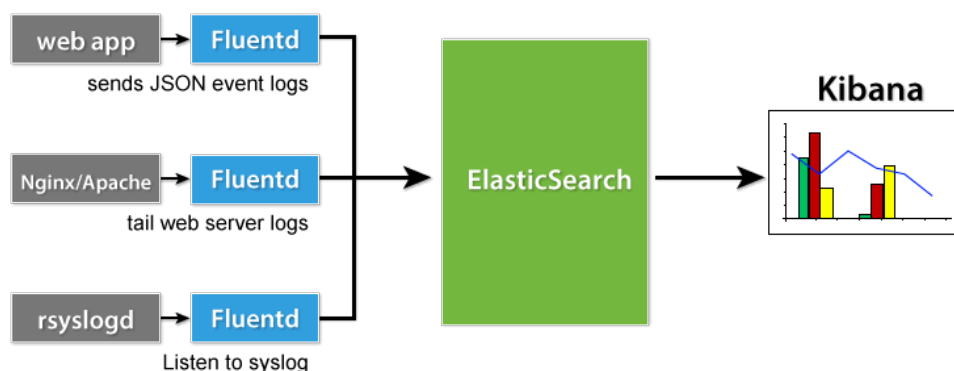


Figura 2.8: L'architettura realizzata con Fluentd, Elastic Search e Kibana

Benché questa architettura presenti caratteristiche e funzionalità aderenti alle esigenze di Codice Plastico, essa richiede una ampia e approfondita conoscenza delle singole applicazioni che la compongono. La frammentazione delle funzionalità su tre diverse applicazioni, inoltre, rende necessaria un'opera di manutenzione attenta e puntuale su tutte le componenti del sistema (aggiornamenti, compatibilità dei software, ecc...). Questo genera un *overhead* di gestione rispetto alla manutenzione di una singola applicazione.

Riepilogo Comparativo

Nella tabella 2.1 sono riportate schematicamente le principali caratteristiche analizzate nelle soluzioni, per la centralizzazione dei log, prese in esame.

	Airbrake	Log.io	Rollbar	Papertrail	Fluentd, ElasticSearch, Kibana
Open source		✓			✓
Persistenza dati	✓		✓	✓	✓
Dati utente strutturati			✓		✓
Aggregazione automatica	✓		✓	✓	✓
SaaS	✓	✓	✓	✓	

Tabella 2.1: Tabella comparativa delle funzionalità

2.2 Una Nuova Applicazione: Mole.io

L'analisi dello stato dell'arte delle tecnologie presenti sul mercato, ha portato CodicePlastico a perseguire la scelta dello sviluppo di una nuova applicazione realizzata internamente all'azienda stessa.

Le motivazioni che hanno determinato questa scelta, sono molteplici:

Business la maggior parte delle soluzioni analizzate è fornita come *Software as a Service* (SaaS). In questo caso sarebbe d'obbligo far adottare tale soluzione ad ogni cliente, perdendo di fatto, il vantaggio di avere un sistema centralizzato di raccolta dei dati.

Riservatezza le soluzioni analizzate che sono fornite come SaaS, prevedono che i messaggi vengano salvati su basi di dati remote e non di proprietà dell'azienda. Questo espone potenzialmente a problematiche di gestione della riservatezza dei dati sensibili che sarebbero più facilmente controllabili utilizzando un sistema *self-hosted*.

Personalizzazione a parte l'ultima soluzione analizzata (Fluentd, Elastic Search e Kibana) la maggior parte dei software presenta difficoltà nella personalizzazione del contenuto dei messaggi, fondamentale per l'azienda, sia per tracciare con precisione problematiche specifiche delle applicazioni in produzione, sia nell'ottica della raccolta dei dati per la Business Intelligence.

Manutenzione limitare gli interventi di manutenzione ad un'unica applicazione proprietaria, di cui, conseguentemente, si ha pieno controllo dello sviluppo.

Estensibilità l'intero sviluppo delle varie componenti dell'applicazione è stato guidato dal concetto di estensibilità. Il sistema risultante dovrà

essere facilmente estendibile con nuove funzionalità e *feature* per adattarsi nel modo più aderente possibile alle esigenze di ogni cliente. Un ulteriore aspetto che ha motivato alla costruzione di una nuova applicazione, è la possibilità di rilasciare gradualmente agli utenti le nuove funzionalità introdotte, man mano, nel sistema (*deploy* graduale).

Sperimentazione tra i valori aziendali fondamentali di CodicePlastico vengono annoverati l'aggiornamento costante del personale e il miglioramento delle *skill* di ogni sviluppatore, anche attraverso pratiche agili come *pair programming*, *randori* e progetti personali. Lo sviluppo di una applicazione con l'ausilio di tecnologie relativamente recenti e innovative nel panorama IT, è in completo accordo con la *vision* aziendale.

Capitolo 3

Metodologie di Sviluppo

Per lo sviluppo del progetto di tesi, sono state adottate tecniche di progettazione e sviluppo del software che provengono dall'ambito delle metodologie di sviluppo agili.

Il modello di sviluppo agile è un insieme di pratiche basate sulla costruzione iterativa e incrementale del software. Il flusso di lavoro è organizzato in cicli di breve durata, che hanno come oggetto l'implementazione di una piccola quantità di *feature*. Questa pratica di sviluppo iterativo, offre un immediato vantaggio: la possibilità di riesaminare il lavoro effettuato al ciclo precedente e decidere, a seconda delle esigenze correnti, se continuare lo sviluppo nella stessa direzione o cambiare radicalmente approccio.

Questa possibilità è fondamentale nell'ottica della buona riuscita di un progetto, poiché è molto frequente che, durante lo sviluppo di applicativi software, i committenti decidano, per vari motivi, di sostituire logiche di funzionamento del software. Lo sviluppo iterativo, quindi, mette in condizione il team di lavoro di proporre soluzioni che si adattano nel tempo alle specifiche, generando così un prodotto strettamente aderente alle aspettative del cliente.

Tra le tecniche operative più frequenti adottate nello sviluppo di un progetto, citiamo:

- pianificazione adattiva
- sviluppo evolutivo
- rilasci frequenti
- *time-boxing*

Lo scopo di queste tecniche è ottenere un modello di sviluppo che possa essere flessibile e adattarsi bene al cambiamento.

Lo Sviluppo Iterativo e Incrementale

Il processo di sviluppo è suddiviso in unità base, con una durata temporale che va da una a quattro settimane. Conseguentemente non è possibile organizzare le attività di sviluppo tenendo conto dei soli obiettivi globali del progetto, ma è necessario ri-contestualizzarli rispetto alla finestra temporale adottata. Gli obiettivi globali vengono così suddivisi in task più piccoli, ciascuno focalizzato su una problematica specifica.

Al termine di ogni unità base, si procede con la successiva, in un processo iterativo. Il risultato di una iterazione dovrebbe essere un prodotto parzialmente funzionante da mostrare al cliente.

Questo approccio minimizza il rischio di portare lo sviluppo fuori dal contesto e permette di avvicinare il cliente al processo di realizzazione del software, rendendolo partecipe dei problemi incontrati e permettendogli di ottenere un prodotto molto aderente alle proprie esigenze.

La Comunicazione Rapida

Le metodologie agili suggeriscono una organizzazione gerarchica, con un ristretto numero di livelli, dei team di sviluppo. Ogni team elegge il proprio *owner* che è responsabile del gruppo di lavoro ed è l'unica interfaccia con i superiori. Anche in questo caso l'esigenza è creare una catena di comunicazione il più corta possibile, in modo da permettere agli sviluppatori di ottenere, in brevissimo tempo, un *feedback* riguardo ai problemi incontrati.

Un altro strumento utile a questo scopo sono gli *stand-up meeting*: riunioni giornaliere brevissime, nelle quali, per salvaguardare la brevità dell'incontro, i partecipanti formano un cerchio stando in piedi. In questi incontri ogni sviluppatore riporta all'*owner* l'elenco dei lavori sul quale è impegnato, stime di tempi per la chiusura delle *feature* in sviluppo, eventuali problematiche incontrate e programmazione delle attività di sviluppo nell'immediato futuro.

Mantenere Alta la Qualità

Le metodologie agili prevedono l'impiego di strumenti software avanzati per supportare le tecniche descritte e facilitare il raggiungimento degli obiettivi prefissati. In [8] e [9] Robert C. Martin descrive alcuni strumenti e varie tecniche utilizzate per perseguire l'obiettivo della qualità del software. Nella trattazione, tra gli altri, sono annoverati:

- *continuous-integration*
- test automatici
- *pair programming*
- *test-driven development* (TDD)

- *design patterns* [10]
- *user stories*

Nelle sezioni seguenti approfondiremo alcune delle tecniche utilizzate durante la progettazione e lo sviluppo del progetto.

3.1 User Story

Le pratiche agili suggeriscono l'utilizzo di *user story* per la descrizione del comportamento del sistema. Una *user story* è una frase, composta utilizzando il linguaggio naturale, che descrive una particolare caratteristica dell'applicazione da implementare.

La struttura della frase da redigere è prefissata e si compone di tre entità fondamentali: *chi*, *cosa* e *perché*. Attraverso questi tre concetti è possibile descrivere il comportamento atteso da un attore nel sistema, umano o software, che esegue una determinata azione, al fine di ottenere un risultato.

Ad esempio, per descrivere la *feature* di autenticazione di un utente nel sistema, si potrebbe procedere nel modo seguente:

Come utente

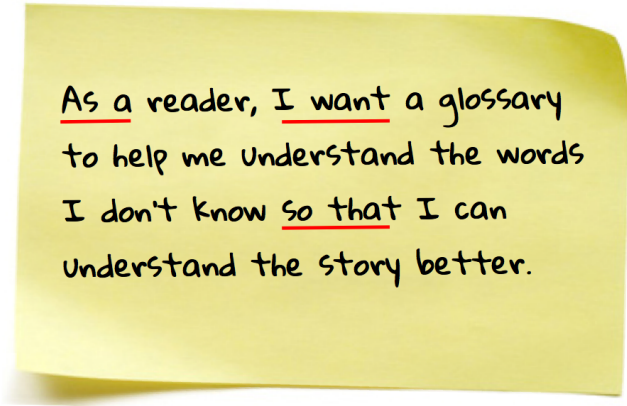
Voglio poter inserire username e password

Al fine di accedere al sistema

Come mostra l'esempio, le frasi devono essere concise, in modo da rappresentare una funzionalità ben definita del sistema.

Le user story sono la prima fase della progettazione e hanno lo scopo, oltre a quello di stilare una lista condivisa di funzionalità da implementare, di definire un ordine di priorità e la stima dei tempi di realizzazione delle stesse. Un metodo semplice ed efficace per stimare le user story è riportare il tempo di realizzazione di ognuna a fianco della descrizione.

La *user story* è tipicamente scritta su un biglietto adesivo simile a quello in figura 3.1, e incollata ad una lavagna che riporta l'insieme dei requisiti del sistema. Un esempio è la *board* in figura 3.2.



As a reader, I want a glossary
to help me understand the words
I don't know so that I can
understand the story better.

Figura 3.1: Un esempio di *user story*

La creazione delle storie, tipicamente, avviene in collaborazione con il cliente. Durante questo processo l'*owner* lo guida, attraverso domande specifiche, alla stesura e formalizzazione delle funzionalità necessarie.

Ogni storia rappresenta una feature ben precisa sviluppabile in un periodo che va, tipicamente, da alcuni giorni a un paio di settimane.

L'operazione fisica di scrittura della user story, aiuta il cliente a concretizzare la sua idea di progetto, visualizzandone i dettagli del funzionamento. Questo è molto utile sia per gli sviluppatori, sia per il cliente stesso. Infatti, spesso, i requisiti richiesti non sono dettagliati e specifici a sufficienza. Come ulteriore vantaggio si ha la costruzione di un linguaggio comune, condiviso, tra cliente e sviluppatori che rende agevole la realizzazione dell'intero progetto e dei futuri feedback.

Nel caso in cui le esigenze di progetto cambino, è molto facile, in questo processo, sia sostituire le user story, sia rendersi conto di quale impatto comporti il cambiamento, sull'intero sistema.

Come sarà descritto nella sezione successiva, l'utilizzo delle user story si lega a quello del TDD, poiché i *test* diverranno la dimostrazione oggettiva

dell'effettiva implementazione e garanzia di correttezza di ogni storia.

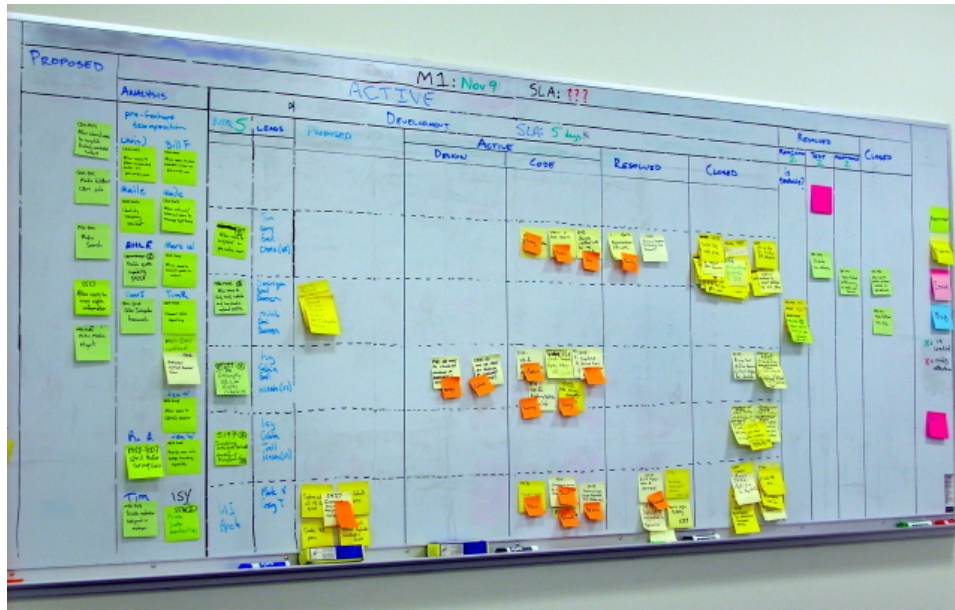


Figura 3.2: Una lavagna con user story

3.2 Il Test Driven Development

Il *test driven development* (TDD) è una tecnica di programmazione che prevede la scrittura di un test, ovvero di una porzione di codice che si occupa di verificare una funzionalità, prima della implementazione della funzionalità stessa.

L'ideazione e la diffusione di questa tecnica di programmazione è stata curata da Kent Beck, che nel suo libro *Test Driven Development: By Example* [11] spiega come applicare il TDD a svariati contesti e dettaglia questa tecnica nelle sue tre fasi fondamentali, definite abitualmente con il termine *red-green-refactor*.

I nomi *red* e *green* delle prime due fasi derivano dalla convenzione, utilizzata dalla maggior parte dei framework di test, di indicare in colore rosso un test fallito e in verde un test andato a buon fine.

Di seguito sono riportati i dettagli di ciascuna fase.

Prima fase: Red

Ogni nuova feature da implementare inizia con la stesura di un nuovo test, questo aiuta il programmatore a definire, a priori, le specifiche funzionali della porzione di sistema che dovrà implementare. La funzionalità descritta dal test può mappare direttamente una *user story* o può essere una porzione di essa.

Poiché la feature oggetto del test non è stata implementata, l'esecuzione del test riporterà esito negativo e, conseguentemente, esso risulterà fallito (rosso).

Questa pratica, a prima vista inutile, ha lo scopo di validare il test stesso. Infatti se il risultato fosse verde, questo significherebbe che esso è superfluo oppure errato.

Seconda fase: Green

Questa fase del processo è quella nella quale si esegue lo sviluppo vero e proprio della nuova funzionalità, in modo da far diventare il test verde (soddisfatto).

L'obiettivo è esclusivamente soddisfare i requisiti quindi lo sviluppatore, tipicamente, cerca di arrivare alla soluzione nel modo più semplice possibile, non curandosi dei dettagli stilistici del codice.

Terza fase: Refactor

L'ultimo passo da eseguire è la rifattorizzazione del codice, cioè la riorganizzazione delle funzionalità realizzate al passo precedente al fine di ottenere una architettura ben congegnata e una struttura chiara.

Una delle tecniche principali per realizzare questo compito è chiamata *Don't Repeat Yourself* (DRY) e consiste nel cercare di rimuovere quanto più possibile il codice duplicato mantenendo le medesime funzionalità.

L'utilizzo di questa tecnica favorisce il riuso e, tipicamente, veicola verso un buon design del software. In questa fase lo sviluppatore ha la massima libertà di sperimentazione, in quanto la correttezza del software rifattorizzato è garantita dai test scritti in precedenza.

Il processo del TDD è iterativo, al termine della fase di refactoring, esso ricomincia da capo, con la stesura di un nuovo test o la riformulazione di test esistenti. Il grafico 3.3 riporta schematicamente l'andamento circolare di questo processo.

La scelta di utilizzare il TDD all'interno di un progetto software garantisce la corretta implementazione delle user story. L'adozione di questa tecnica è, inoltre, facilitata dalla esistenza di innumerevoli framework di test per la maggior parte dei linguaggi di programmazione esistenti.

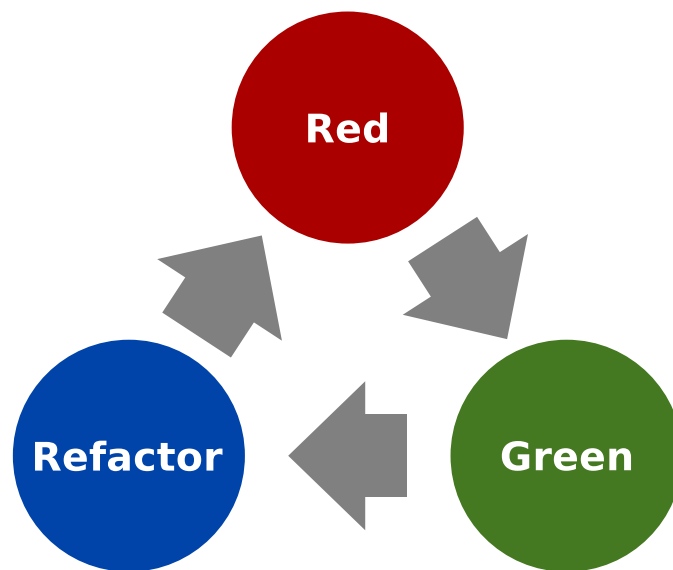


Figura 3.3: Il flusso di lavoro circolare del TDD

L'utilizzo del TDD fornisce agli sviluppatori svariati vantaggi, alcuni dei quali sono:

- riduzione drastica della necessità di un *debugger* per la ricerca degli errori;
- acquisizione di fiducia nella sperimentazione, poiché l'impatto dell'adozione ogni soluzione è sempre validato dai test e quantificabile;
- documentazione del codice per gli altri sviluppatori del team, i test forniscono infatti esempi di utilizzo delle diverse funzionalità implementate;
- tracciamento dello stato di avanzamento dei lavori;
- modularizzazione della struttura del progetto, tramite la suddivisione delle feature in unità indipendenti e cooperanti tra loro;
- *focus* su una singola funzionalità;

Capitolo 4

Tecnologie Utilizzate

In questo capitolo approfondiremo i dettagli delle tecnologie utilizzate per realizzare Mole.io. Illustreremo, per ognuna di esse, le motivazioni che ci hanno spinto alla scelta di particolari soluzioni software e le problematiche incontrate durante il loro utilizzo.

Mole.io è stato interamente sviluppato utilizzando il linguaggio JavaScript. L'utilizzo di questa tecnologia è abbastanza comune all'interno delle pagine web e si presta bene all'utilizzo *client side*, meno comune è invece la sua applicazione nella parte *server*. La piattaforma Node.js permette di utilizzare JavaScript per sviluppare la parte *backend* delle applicazioni. Rendere uniforme il linguaggio utilizzato permette facilitare lo sviluppo e la fruibilità del progetto da parte degli sviluppatori. Per poter lavorare al progetto è infatti richiesta solo la conoscenza di JavaScript e non di altri linguaggi, questo è un ottimo requisito quando si pensa ad un team di sviluppo in espansione.

Quella del paragrafo precedente è solo una delle motivazioni che ci hanno spinto a scegliere Node.js come tecnologia di sviluppo, iniziamo quindi a vedere in dettaglio questa tecnologia.

4.1 Node.js

La *homepage* del sito ufficiale di Node.js [12] fornisce una sintetica ma precisa descrizione di questa tecnologia, partiremo proprio da essa per illustrarne le peculiarità.



Figura 4.1: Il logo ufficiale di Node.js

La descrizione di Node.js riportata sul sito recita:

Node.js is a platform built on Chrome's JavaScript runtime for easily building fast, scalable network applications. Node.js uses an event-driven, non-blocking I/O model that makes it lightweight and efficient, perfect for data-intensive real-time applications that run across distributed devices.

Scopriamo immediatamente che Node.js è una *piattaforma*. L'utilizzo di questo termine mette l'accento su un aspetto fondamentale di questa tecnologia: fornire un ambiente nel quale le applicazioni sviluppate possano funzionare con il supporto di librerie di sistema fornite da Node.js stesso.

La piattaforma Node.js utilizza JavaScript come linguaggio di sviluppo. Per farlo si avvale del potente interprete *V8* presente all'interno del browser *Chrome* di *Google*. L'utilizzo di un linguaggio altamente popolare e di una base solida come quella fornita dal popolare *browser* permettono di costruire applicazioni in modo semplice e veloce.

L'ultimo importante concetto, che leggiamo dalla prima frase della descrizione, è che Node.js è principalmente orientato allo sviluppo di applicazioni che lavorano con la rete. Per sua natura, la piattaforma ci aiuta a fare in modo che esse siano scalabili.

La seconda parte della descrizione spiega sinteticamente alcune caratteristiche peculiari di Node.js e ne definisce meglio il contesto applicativo.

L'intera piattaforma è centrata sul concetto di *evento*. Si dice evento un messaggio che viene scatenato in un determinato istante dell'elaborazione e che successivamente è catturato e gestito dalle componenti del sistema che sono preposte alla gestione di quell'evento specifico.

Il sistema ad eventi viene utilizzato da Node.js congiuntamente ad una gestione non bloccante delle operazioni di *Input/Output*. Questo significa che una operazione potenzialmente lunga, come ad esempio la comunicazione con il *filesystem* o con i dispositivi di *rete* non bloccano il flusso di esecuzione del programma principale. Dopo aver richiesto ad altri attori del sistema il dato di cui necessita, il programma continua il suo normale flusso di esecuzione e verrà informato, utilizzando un evento, quando il dato richiesto sarà disponibile. Questa gestione non bloccante delle operazioni di *I/O* è chiamata *I/O* asincrono.

Questa modalità di gestione delle operazioni non strettamente legate alla logica applicativa, permette a Node.js di essere molto efficiente se utilizzato per la realizzazione di applicazioni che elaborano grandi quantità di dati ma devono rimanere *reattive* nei confronti di nuove richieste di elaborazione.

4.1.1 Le operazioni asincrone

Abbiamo introdotto il concetto di I/O asincrono, di seguito illustreremo come Node.js riesca a gestirlo utilizzando una quantità limitata di risorse di sistema.

I componenti *hardware* di un sistema elaborano dati con velocità differenti. La rapidità di ogni componente è fortemente legata al modo nel quale questo è stato realizzato. La tabella seguente riporta un elenco di componenti e le relative velocità indicative riferite ad un singolo ciclo di *CPU*.

Componente	Numero di cicli
CPU	1
Cache di livello 1	3
Cache di livello 2	14
RAM	250
Hard Disk	41.000.000
Dispositivo di rete	240.000.000

Guardando la tabella si nota immediatamente come i dispositivi di *I/O* siano ordini di grandezza più lenti rispetto ai dispositivi di elaborazione delle informazioni. Se il flusso del programma dovesse aspettare in modo sincrono ogni singola operazione di lettura o scrittura su disco, ad esempio, esso perderebbe la possibilità di eseguire circa quaranta milioni di operazioni di calcolo, con un conseguente degrado delle performances del software.

Una tecnica comune per affrontare il problema dell'*I/O* è l'utilizzo di *threads*. Un thread è spesso definito con il termine sotto-processo leggero, in effetti esso condivide codice e risorse di sistema con altri threads appartenenti allo stesso processo padre.

Un thread permette al processo di parallelizzare l'esecuzione di una parte del suo flusso di lavoro, ma al tempo stesso sono necessarie risorse di sistema per creare il thread stesso e per distruggerlo. Node.js ovvia a questo problema utilizzando un *thread pool*, cioè un insieme di threads che fungono da *worker* per il processo. Quando è necessario eseguire un particolare task, esso viene sottoposto al thread pool, di conseguenza viene assegnato ad un worker, esso lo esegue e al termine del lavoro comunica l'esito dell'elaborazione al chiamante tramite una funzione detta *callback*.

L'utilizzo di un thread pool comporta un incremento di prestazioni da parte delle applicazioni che lo sfruttano: i thread pool ottimizzano l'utilizzo della memoria e del processore, diminuendo l'overhead di gestione dei thread.

4.1.2 Le callback

Una delle principali difficoltà quando si utilizza Node.js per la prima volta è la gestione delle callback. Di seguito è riportato un frammento di codice Node.js che esegue il salvataggio di un utente su Database.

```
function save(user) {  
  console.log('saving user into db');  
  db.insert(user, function(err, user) {  
    console.log('user successfully saved!');  
  });  
  console.log('all we need is just a little patience...');  
}
```

la funzione `db.insert()` è asincrona, ed accetta due parametri: il dato da salvare e la callback da eseguire una volta eseguito l'inserimento nel Database. L'esecuzione di questo codice produce un output in console simile a quello mostrato qui sotto.

```
saving user into db
all we need is just a little patience...
user successfully saved!
```

Se la funzione `db.insert()` non fosse asincrona otterremmo un output come quello che segue.

```
saving user into db
user successfully saved!
all we need is just a little patience...
```

Questo piccolo esempio, lascia intuire un potenziale problema nella gestione delle callback: se, ad esempio, dovessimo assegnare alcuni privilegi di *default* a tutti i nuovi utenti inseriti nel sistema, potremmo scrivere il seguente codice.

```
function save(user) {
  console.log('saving user into db');
  db.insert(user, function(err, user) {
    db.insert(grants, user, function(err, grants) {
      console.log('user successfully saved!');
    });
  });
  console.log('all we need is just a little patience...');
}
```

Ora il problema è evidente: se i dati ottenuti in modo asincrono sono necessari per eseguire altre procedure, essi vanno gestiti internamente alla callback stessa. Se utilizzate in modo improprio, le callback, portano al cosiddetto *callback-hell*, ovvero un codice nel quale il lettore non riesce più a seguire il flusso logico dell'applicazione.

Fortunatamente JavaScript ci fornisce una soluzione semplice al problema, ma che richiede allo sviluppatore disciplina nello scrivere il codice sorgente. L'esempio precedente, infatti potrebbe essere riscritto come segue.

```
function onGrantsSaved(err, grants) {
  console.log('user successfully saved!');
}

function onUserSaved(err, user) {
  db.insert(grants, user, onGrantsSaved);
}

function save(user) {
  console.log('saving user into db');
  db.insert(user, onUserSaved);
  console.log('all we need is just a little patience...');
}
```

Il codice è ora molto più leggibile, semplicemente evitando la dichiarazione di funzioni anonime in linea.

4.1.3 La Storia

4.1.4 Npm e Moduli

L'organizzazione del codice è un aspetto fondamentale dello sviluppo. Il *Single Responsibility Principle* (SRP) è uno dei principi di design del software più importanti. Applicato alla programmazione ad oggetti, esso sostiene che ogni oggetto deve essere responsabile di un singolo aspetto del comportamento del sistema.

Node.js incarna questo principio nelle fondamenta della sua struttura, infatti permette di organizzare il codice in unità indipendenti tra loro chiamate *moduli*. Ogni modulo nella piattaforma può decidere quali dati gestire al suo interno e quali esporre all'esterno. Si creano in questo modo delle *black box* che funzionano tanto meglio quanto il loro compito è specifico.

Poco dopo la nascita di Node.js la comunità di sviluppatori che lo utilizzava ha deciso di arricchire questa piattaforma con un *tool* ormai insostituibile: *Node Package Manager* (NPM).



Figura 4.2: Il logo di NPM

NPM [13] è un tool, utilizzabile da linea di comando, che si occupa della gestione dei moduli e delle loro dipendenze, per farlo utilizza un *repository* in rete nel quale sono registrati tutti i moduli pubblici sviluppati dai vari *contributor* della *community* Node.js.

Abbiamo introdotto il concetto di *dipendenza* tra moduli. Un modulo

si dice dipendente da un altro quando necessita di quest'ultimo per eseguire il suo compito. NPM impone che ogni modulo sia descritto dal file `Package.json` che ne riporta le informazioni principali ed elenca gli altri moduli dai quali dipende. Di seguito riportiamo un esempio di `Package.json` per un modulo chiamato `mole` che fa parte del sistema realizzato per questa tesi.

```
{
  "name": "mole",
  "version": "0.0.1",
  "author": "Federico Gandellini",
  "description": "whisper collector and denormalizer",
  "private": true,
  "scripts": {
    "start": "node mole.js"
  },
  "engines": {
    "node": ">=0.8.0",
    "npm": ">=1.2.0"
  },
  "dependencies": {
    "express": "3.3.4",
    "underscore": "~1.5.2",
    "mongo-make-url": "0.0.1",
    "mongoskin": "~0.6.0",
    "mongodb": "~1.3.19",
    "require-all": "0.0.8",
    "rabbit.js": "~0.3.1"
  }
}
```

```
  },  
  "devDependencies": {  
    "grunt-contrib-jshint": "~0.6.4",  
    "grunt": "~0.4.1",  
    "grunt-mocha-cli": "~1.2.1",  
    "grunt-contrib-watch": "~0.5.3",  
    "mocha": "~1.13.0",  
    "should": "~1.3.0",  
    "supertest": "~0.8.0",  
    "Faker": "~0.5.11"  
  }  
}
```

Nella prima parte del file possiamo trovare i dati principali del modulo, come il suo nome, l'autore, la versione, e una descrizione. Seguono un paio di parametri che definiscono le regole di pubblicazione, il comando per lanciare questo modulo e le versioni richieste della piattaforma Node.js e di NPM.

Le due sezioni seguenti nel file elencano le dipendenze, la prima indica i moduli che devono essere presenti perché esso possa essere messo *in produzione*, le altre sono dipendenze che sono richieste esclusivamente durante lo sviluppo o il *testing* del modulo stesso. Come si può notare, nel file, non sono presenti solo i nomi, ma anche le versioni richieste degli altri moduli.

Uno dei comandi più utilizzati di NPM è `npm install`, esso legge il file `Package.json` presente nella *directory* corrente e scarica dalla rete tutti i pacchetti richiesti alle rispettive versioni e permette all'utilizzatore del modulo di essere immediatamente operativo.

Alcuni moduli utilizzati

Per poter utilizzare le funzionalità fornite da un modulo aggiuntivo, Node.js fornisce un comando che permette di importarlo dall'esterno. Questo comando è `require()` e si utilizza passando come parametro il nome del modulo da caricare.

```
var express = require('express');
```

Una volta eseguito il comando, la variabile `express` conterrà il modulo appena caricato e sarà possibile utilizzarne le funzionalità.

Illustriamo ora alcuni dei moduli utilizzati per realizzare la nostra applicazione.

express È uno dei moduli più importanti: permette di creare facilmente un *server web* in grado di rispondere a richieste provenienti dai *client*. Utilizza un sistema di *rotte* per legare l'azione da eseguire alla richiesta HTTP in arrivo. Express eredita da *Connect* una funzionalità chiave per il suo funzionamento, i *middleware*. In Connect, un middleware è una funzione che filtra tutte le richieste in ingresso e le risposte in uscita e le restituisce rielaborate. I middleware sono costruiti in modo da essere accodati l'uno con l'altro, dando la possibilità di costruire filtri molto complessi. Illustreremo la configurazione delle rotte e dei middleware di express nella sezione 5.1.

passport Il compito di questo modulo è gestire l'autenticazione degli utenti. Passport fornisce un comodo middleware per express, con il quale è possibile verificare i dati di autenticazione dell'utente quando questo voglia accedere a specifiche rotte. Più avanti, nella sezione 5.2, approfondiremo questo tema.

mongoose e mongoskin Sono i due principali *driver* Node.js per MongoDB, il sistema per l'archiviazione dei dati che abbiamo utilizzato nell'applicazione. Nella sezione 4.3 vedremo nel dettaglio questo database documentale e illustreremo come è possibile accedere alla sua interfaccia utilizzando Node.js.

require-all Permette di caricare tutti i moduli presenti in una directory. Questo modulo ci è stato molto utile per garantire e semplificare l'estensibilità del sistema. Nella sezione 5.1.1 vedremo nel dettaglio come abbiamo sfruttato questa semplice ma potente funzionalità.

mocha Questo modulo non fornisce funzionalità vere e proprie da spendere in produzione, bensì un insieme preziosissimo di *tool* per poter testare la propria applicazione Node.js. Abbiamo largamente utilizzato questo strumento durante la realizzazione del sistema per renderlo, quanto più possibile, *bug-free*.

4.2 RabbitMQ

Quando si progetta una applicazione web che offre un servizio, bisogna sempre porre molta attenzione a non introdurre nel sistema i cosiddetti *colli di bottiglia*, cioè componenti dell'architettura che non riescono a sopportare il carico delle richieste in arrivo dagli utenti.

Una tecnica piuttosto efficace per evitare i colli di bottiglia, consiste nel realizzare un *disaccoppiamento* delle varie componenti presenti nel sistema. Una applicazione ben disaccoppiata è una applicazione nella quale ogni singolo componente svolge una funzione specifica e comunica con le altre parti del sistema secondo protocolli predefiniti. Disaccoppiare quindi non è sempre facile, perché è necessario costruire infrastrutture che permettano alle varie parti, ormai slegate, di comunicare tra loro.

RabbitMQ [14] è un software che permette di realizzare, configurare e monitorare complessi sistemi di code di messaggi. Il suo utilizzo permette di realizzare una infrastruttura nella quale le diverse parti di un sistema possano comunicare tra loro scambiandosi messaggi attraverso le code utilizzando un protocollo determinato dallo sviluppatore.



Figura 4.3: Il logo di RabbitMQ

L'utilizzo di RabbitMQ fornisce un vantaggio evidente in termini di flessibilità: è possibile agganciare o rimuovere parti da un sistema attivo senza comprometterne l'intero funzionamento.

La flessibilità non è l'unico vantaggio, infatti RabbitMQ permette di ottenere architetture perfettamente gestibili su sistemi di tipo *Platform as a Service* (PaaS), sui quali si può decidere di aumentare o diminuire dinamicamente le risorse fornite ad un sistema in produzione in accordo con il numero di richieste utente da soddisfare.

RabbitMQ permette di costruire *pattern* di comunicazione tra processi, i più utilizzati sono quelli riportati di seguito.

Work Queues

L'idea alla base questo tipo di configurazione, chiamata anche *Task Queues*, consiste nell'evitare picchi nel carico di lavoro e risorse del sistema, dovuti allo svolgimento di una operazione e all'attesa del suo completamento. Per ovviare a questo problema si configura un sistema nel quale esiste un produttore di dati e uno o più consumatori.

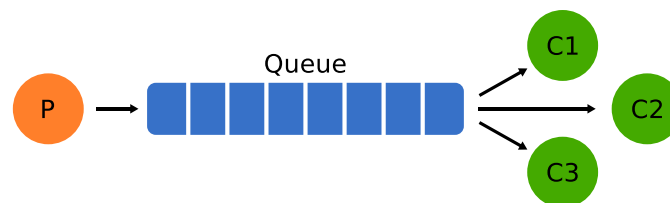


Figura 4.4: Work Queues in RabbitMQ

In figura 4.4 vediamo una rappresentazione schematica di questo tipo di configurazione. All'arrivo di una richiesta di elaborazione, il produttore (P) invia un messaggio sulla coda. I consumatori (C1, C2 e C3) in attesa estraggono un messaggio dalla coda e lo elaborano.

Nella sua configurazione base, la coda esegue un *load-balancing* dei messaggi, questo significa che fornisce esattamente lo stesso quantitativo di

messaggi, e quindi di carico di lavoro, ad ogni consumatore.

RabbitMQ permette di variare il numero di consumatori in ascolto sulla coda dinamicamente, a *runtime*. Non appena un nuovo consumatore si registra per la ricezione dei messaggi, il carico di lavoro viene automaticamente ricalcolato in modo da essere costante per tutti i nodi.

Il load-balancing automatico e la possibilità di sottoscrivere consumatori a runtime diventano funzionalità molto importanti in presenza di sistemi attivi su PaaS. All'aumentare delle richieste, infatti, il sistema potrebbe in automatico attivare nuovi consumatori e sottoscriverli, ottenendo in questo modo un sistema altamente scalabile.

Publish/Subscribe

Questa configurazione si ispira al concetto di *abbonamento*, l'idea di base infatti è poter inviare il medesimo messaggio a più destinatari.

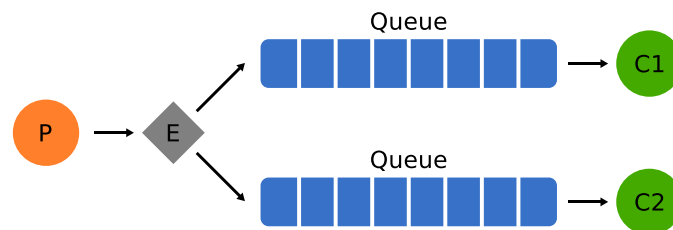


Figura 4.5: Publish/Subscribe in RabbitMQ

La figura 4.5 mostra una rappresentazione del sistema Publish/Subscribe. A differenza del modello *Work Queues*, qui è stato aggiunto un nuovo attore: un *exchange* (E) il cui compito è duplicare i messaggi e inviarli a tutte le code ad esso connesse.

In questo caso il produttore di messaggi (P) invia un messaggio ad un exchange (E), il quale lo duplica e lo invia a tutte le code ad esso connesse.

I consumatori in ascolto (C1 e C2), riceveranno lo stesso messaggio.

Nella sezione 5.1 vedremo nel dettaglio la configurazione di RabbitMQ utilizzata per realizzare Mole.io e quali tecniche abbiamo messo in atto per permettere al sistema di essere installato su una piattaforma di tipo PaaS.

4.3 MongoDB

Il salvataggio dei dati è una operazione critica in moltissimi sistemi software. La scelta di quale database utilizzare per salvare le informazioni è altrettanto delicata e va ponderata alla luce di vari punti di vista. Per la nostra applicazione abbiamo scelto di utilizzare MongoDB. Di seguito cercheremo di illustrare le principali caratteristiche di questo database e le motivazioni che ci hanno spinto a sceglierlo per la nostra applicazione.



Figura 4.6: Il logo di MongoDB

I database relazionali (RDBMS), per loro natura, espongono il loro contenuto in un formato tabellare e utilizzano concetti come righe e colonne per fornire l'accesso ai dati o porzioni di essi. In questo tipo di database lo *schema* dei dati, cioè la loro struttura è ben definita e va studiata in fase di *design* della base di dati, essi si definiscono infatti database *schema-full*. Negli RDBMS la modifica del formato di un dato a sistema avviato è una operazione delicata. Essa infatti deve tenere conto dei dati già presenti all'interno del sistema e deve aggiornarli coerentemente con la nuova struttura assunta dalla tabella che li contiene.

MongoDB è un database *schema-less*, questo significa che la struttura di un dato non è definita a priori, ma soprattutto che essa può variare nel tempo senza richiedere aggiornamenti ai dati già presenti nella base dati. MongoDB utilizza infatti un modello documentale per il salvataggio dei

dati, essi salvati internamente come oggetti *BSON* e forniti all'esterno sotto forma di oggetti *JSON*.

Per comprendere i concetti alla base di MongoDB è possibile realizzare alcune similitudini tra concetti proposti da questa base dati e concetti presenti negli RDBMS. La tabella seguente mostra alcune di queste similitudini.

RDBMS	MongoDB
table	collection
tuple	document
column	field

Le *collection* sono insiemi di *document*, i quali, a loro volta, contengono vari *field* che rappresentano le chiavi per l'accesso ai dati veri e propri: i *value*.

Un documento BSON può contenere *field* di vario tipo: interi, stringhe, *array*, dati binari, oppure altri documenti (*embedded document*). Come abbiamo anticipato, in MongoDB, lo schema dei documenti non è fisso, questo significa che nella stessa *collection* potremo trovare *document* con struttura differente.

Altre funzionalità significative di MongoDB sono la possibilità di eseguire aggiornamenti atomici dei dati, la ricerca *full-text* all'interno dei documenti e degli *embedded document*, la possibilità di creare indici sui dati e indici di tipo geospaziale.

La possibilità di salvare dati aventi una struttura variabile è stato il motivo principale che ci ha spinto ad utilizzare questo tipo di database, nel capitolo 5.1.2 vedremo come questa funzionalità ci ha permesso di creare un sistema estremamente flessibile.

I creatori di MongoDB hanno fatto in modo che il database da loro realizzato realizzasse due caratteristiche fondamentali, che lo rendono il candidato

ideale per l'installazione in ambienti PaaS: l'alta accessibilità e la scalabilità.

Nella sezione seguente vedremo come MongoDB implementa questi due concetti e come essi si possono utilizzare sia per fronteggiare l'aumento di richieste da parte degli utenti, sia per rendere il sistema resistente a problemi di malfunzionamento dei server sui quali MongoDB viene installato.

Nella sezione 5.1 vedremo inoltre come queste caratteristiche hanno reso MongoDB il candidato ideale ad essere utilizzato all'interno del nostro sistema. Nella configurazione finale, infatti, esso verrà installato proprio su un servizio di tipo PaaS.

4.3.1 Fronteggiare le Richieste

Come abbiamo anticipato, MongoDB è un database documentale che garantisce alte *performance*, alta accessibilità e permette facilmente di scalare la sua struttura per fronteggiare le richieste. Vediamo brevemente come MongoDB ottiene ognuna di queste funzionalità:

Database documentale I documenti contenuti in MongoDB (oggetti) mappano molto bene gli oggetti e le strutture dati fornite dai principali linguaggi di programmazione, rendendo quasi inutile la necessità di un software di traduzione tra strutture dati utilizzate nel software e la loro rappresentazione nel database. Tipicamente i software di *Object-Relational Mapping* (ORM) sono necessari in presenza di database relazionali. Il vantaggio di avere gli *embedded document*, inoltre, permette di ridurre il numero di operazioni di *join* sui dati, rendendo superflua una delle caratteristiche fondamentali dei RDBMS. Con MongoDB diventa quindi molto semplice far *evolvere* le proprie strutture dati, rendendo lo sviluppo dell'applicazione molto più flessibile.

Alte *performance* La possibilità di inserire documenti in documenti, garantisce scritture veloci e la definizione degli indici può includere chiavi presenti negli *embedded documents*, in questo modo è possibile ottenere tempi di risposta del sistema molto contenuti.

Alta accessibilità Questa proprietà, chiamata tecnicamente, *High Availability* (HA), è realizzata con server replicati e organizzati in *cluster*, detti *replica-set*, nei quali è possibile identificare un *master* e altri *slave*. Il master riceve le richieste e le smista sugli slave nel cluster. In caso di problemi al server master, MongoDB esegue una elezione automatica del nuovo master tra gli slave rimanenti.

Facile scalabilità Lo *sharding* è la possibilità di suddividere una collection su più server in modo automatico. Questa caratteristica rende MongoDB altamente indicato per essere installato su piattaforme di tipo PaaS, all'aumentare dei dati presenti nel database, infatti, è sufficiente aumentare il numero di server a disposizione per accogliere più informazioni. Dal punto di vista dell'applicazione che utilizza questi dati, questa operazione è trasparente. In questo modo il numero di server necessari aumenta linearmente all'aumentare della quantità di dati salvata. È possibile aggiungere server in modo dinamico, senza arrestare il sistema, questa funzionalità è particolarmente importante quando MongoDB è utilizzato per contenere dati di applicazioni web che non ammettono momenti di *downtime*.

Oltre alla flessibilità offerta dal modello documentale, MongoDB include le funzionalità comuni degli RDBMS, quali indici, aggregazioni, *query*, ordinamenti, aggiornamenti di dati aggregati e *upsert*, cioè aggiornamento di un dato se già esistente o creazione di un nuovo dato.

Gli sviluppatori di MongoDB hanno cercato di realizzare un database che fosse semplice da installare e mantenere, infatti la filosofia alla base di MongoDB è fare la cosa giusta. Questo significa che il sistema cerca di adattarsi nel miglior modo possibile alla configurazione dell'*hardware* che lo ospita e fornisce un insieme limitato di parametri di configurazione, mantenendo così una interfaccia *user-friendly* verso gli amministratori e permettendo agli sviluppatori di concentrarsi sulle logiche applicative invece di occuparsi della configurazione del database.

Sebbene MongoDB supporti configurazioni *standalone* (o *single-instance*), la configurazione comune di questo database in produzione è quella distribuita. Combinando le funzionalità di *replica-set* e *sharding* è possibile ottenere

alti livelli di ridondanza per grandi basi di dati in modo completamente trasparente per l'applicazione.

4.4 AngularJS e Altre Tecnologie di Frontend

Why AngularJS?

HTML is great for declaring static documents, but it falters when we try to use it for declaring dynamic views in web-applications. AngularJS lets you extend HTML vocabulary for your application. The resulting environment is extraordinarily expressive, readable, and quick to develop.

Alternatives

Other frameworks deal with HTML's shortcomings by either abstracting away HTML, CSS, and/or JavaScript or by providing an imperative way for manipulating the DOM. Neither of these address the root problem that HTML was not designed for dynamic views.

Extensibility

AngularJS is a toolset for building the framework most suited to your application development. It is fully extensible and works well with other libraries. Every feature can be modified or replaced to suit your unique development workflow and feature needs. Read on to find out how.

4.4.1 Gestione delle Dipendenze

4.5 Strumenti per il Deploy

Capitolo 5

Mole.io

In questo capitolo descriveremo nel dettaglio Mole.io: un nuovo sistema per la gestione centralizzata dei log, realizzato come progetto di questa tesi.

Iniziamo immediatamente con la doverosa spiegazione dell'origine del nome di questo software.

Mole è una parola inglese che significa *talpa*. Nel gergo dello spionaggio, la talpa, è un infiltrato che viene inserito in un sistema avversario e cattura informazioni che riferisce all'*intelligence* della sua fazione.

Il nostro sistema, si comporta esattamente come un infiltrato: passa informazioni del sistema nel quale viene inserito, le applicazioni da monitorare, alla sua organizzazione, gli sviluppatori.

Ogni talpa che si rispetti ha alcuni contatti all'interno del sistema che gli riportano le informazioni rilevanti. Abbiamo chiamato questi contatti *mole-contacts* e le soffiate da loro riferite *whispers*.

Nel nostro sistema, i *mole-contacts* sono moduli software che risiedono all'interno dell'applicazione da monitorare, catturano le situazioni significative per il software nel quale operano ed inviano degli *whisper* ad un server chiamato *mole*.

Nell'immaginario collettivo l'infiltrato è una persona ben vestita, con un abito elegante, giacca, cravatta e cappello. Anche il nostro software, in un certo senso, è ben vestito, infatti possiede una interfaccia grafica realizzata per monitorare le applicazioni ed organizzare gli whisper in arrivo. Questo componente si chiama *mole-suit*.

Il progetto al momento è un prototipo, ma il è destinato ad essere ultimato per diventare un servizio vero e proprio fornito via web. in figura 5.1 è riportato il logo proposto per Mole.io.



Figura 5.1: Il logo di Mole.io

Nella sezione seguente vedremo quali sono le funzionalità specifiche di ogni componente in Mole.io, con particolare attenzione al modo nel quale tali componenti scambiano dati tra loro.

5.1 Architettura del Sistema

Come abbiamo anticipato nell'introduzione di questo capitolo, Mole.io è organizzato in diverse componenti, ognuna delle quali possiede un compito specifico e interagisce con altre.

Nell'architettura di Mole.io si possono innanzitutto individuare due macro-componenti principali. Nelle sezioni seguenti ne illustreremo la struttura.

Insertion

Questo *layer* dell'applicazione si occupa dell'inserimento dei dati nel sistema e si avvale di diversi moduli per fare in modo che gli whisper vengano salvati, aggregati e organizzati nel database.

Nello schema ?? è riportata l'architettura del layer di insertion, essa è composta da diversi moduli:

mole-contacts

la porzione di sistema che si occupa dell'inserimento dei dati provenienti dall'esterno, è composta a sua volta dai *mole-contacts*, da *mole* e dai *denormalizers*.

insertion è la porzione di sistema che si occupa dell'inserimento dei dati provenienti dall'esterno, è composta a sua volta dai *mole-contacts*, da *mole* e dai *denormalizers*.

presentation si occupa dell'estrazione dei dati dal sistema e della loro presentazione all'utente.

5.1.1 CQRS ed Estensibilità

5.1.2 Mole

I Denormalizzatori

5.1.3 Mole Suit

I Plugin e gli Widget

5.2 Autenticazione degli Utenti

5.3 Scalabilità e Affidabilità

5.4 Problematiche di Sviluppo

Capitolo 6

Configurazioni e Benchmark

problemi con i benchmark - dipendi dalla rete su cui sei - nostro client fatto con node - perché non l'abbiamo usato - come sono stati fatti i benchmark - specifiche del sistema VM, ram, hdd, ... - risultati ottenuti

non ha senso testare mole (server) per la parte rabbit (spiegare) il vero collo di bottiglia è il db abbiamo testato con configurazioni di db differenti

Conclusioni e Sviluppi Futuri

Bibliografia

- [1] Engineering Ingegneria Informatica S.p.A. Spagobi website, 2006.
- [2] Inc Rackspace US. Airbrake website.
- [3] Michael Smathers. Log.io website, 2011.
- [4] Brian Rue. Rollbar website, 2004.
- [5] Papertrail Inc. Papertrail website, 2010.
- [6] Inc. Treasure Data. Fluentd website, 2011.
- [7] Elasticsearch. Elasticsearch website, 2009.
- [8] Robert C. Martin. *Clean Code: A Handbook of Agile Software Craftsmanship*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1 edition, 2008.
- [9] Robert C. Martin. *The Clean Coder: A Code of Conduct for Professional Programmers*. Prentice Hall Press, Upper Saddle River, NJ, USA, 1st edition, 2011.
- [10] Elisabeth Freeman, Eric Freeman, Bert Bates, and Kathy Sierra. *Head First Design Patterns*. O' Reilly & Associates, Inc., 2004.
- [11] Beck. *Test Driven Development: By Example*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2002.

- [12] Inc Joyent. Node.js website, 2009.
- [13] Inc Joyent. Node package manager (npm) website, 2010.
- [14] Rabbit Technologies Limited. Rabbitmq website, 2006.