

Università degli Studi di Modena e Reggio Emilia

Dipartimento di Scienze Fisiche, Informatiche e Matematiche

Corso di laurea in Informatica

Engim report service

Relatore:

Prof. Claudia Canali

Candidato:

Dumitru Frunza

Anno accademico 2021/2022

Elenco delle figure

1	Architettura	2
---	------------------------	---

Indice

1	Introduzione	1
2	Engim Srl	1
2.1	L'azienda	1
2.2	Infrastruttura	2
2.3	Microservizi	3
2.4	Sistema automatico di generazione di report	4
3	Requisiti del progetto	5
3.1	Descrizione	5
3.2	Sicurezza	5
3.3	Correttezza dei dati	6
3.4	Problemi interni del server	7
3.5	Presupposti e dipendenze	7
3.6	Features	7
4	Sviluppo	8
4.1	Workflow	8
4.2	Creazione di un file e salvataggio su S3	8
4.3	Stampa di un file pdf da un json	8
5	Implementazione	9
6	Applicazione e performance	9

7	Conclusione	9
8	Appendici	9
9	Bibliografia	9

1 Introduzione

2 Engim Srl

2.1 L'azienda

Engim è una società che si occupa di creare soluzioni tecnologiche in ambito ICT, telecomunicazioni, sistemi di gestione e mobilità. Da oltre 10 anni operano nel mercato della tracciabilità di flotte e attività e della sicurezza dei lavoratori in solitario.

ServizioGPS è il noleggio di tracker gps per veicoli lavoratori. Una prevalente parte dei clienti sono comuni che, tramite i prodotti Engim, tracciano il percorso delle macchine spazzaneve e spargisale. I tracker possono essere prodotti fisici oppure un app per smartphone. A loro volta i prodotti fisici si dividono in fissi e mobili. Il servizio include anche un gestionale per poter visualizzare, modificare o archiviare i propri dati.

Twicetouch è noleggio di dispositivi di sicurezza individuale. Il prodotto tutela i lavoratori in solitario mandando una segnalazione in caso di emergenza.

Esistono due tipi di rilevazione:

- caduta: l'accelerometro del dispositivo rileva un urto pericoloso
- assenza di movimento: il lavoratore non si è mosso per un lasso prolungato di tempo, quindi si presume che possa essere incosciente

Similmente a servizioGPS è possibile noleggiare un dispositivo fisico (badge) oppure l'app per android. In entrambi i casi è possibile impostare i numeri in caso di emergenza, che riceveranno una chiamata e un messaggio SMS.

2.2 Infrastruttura

Le tecnologie usate per servizioGPS sono le seguenti:

- Ruby on rails full stack
- Mariadb e Redis come database
- Python come back end di supporto
- Java per il prodotti app

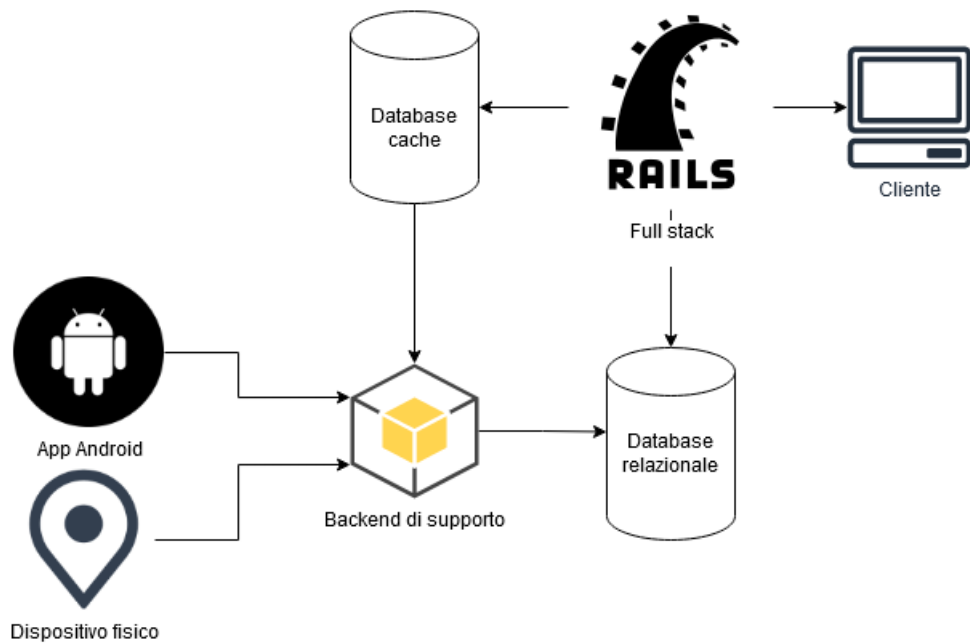


Figura 1: Architettura

Ruby on Rails è usato per front-end e gran parte del backend di servizioGPS. Il fetcher invece si dedica esclusivamente alla elaborazione di coordinate gps,

per poter alleviare il carico di lavoro da Rails. Allo stesso scopo il servizio si appoggia su molteplici server. Uno di questi server è Amazon Web Services, un servizio di cloud computing che oltre al noleggio di un server tradizionale rende possibile anche un architettura a microservizi.

2.3 Microservizi

Un servizio è un processo che: esegue specifiche operazioni autonomamente, risponde a eventi oppure rimane in attesa di una richiesta. Nel caso in cui queste operazioni vengano svolte continuamente, il servizio è estremamente vantaggioso. Non è però necessario che il servizio sia sempre in esecuzione se viene usato in maniera discontinua o per brevi periodi di tempo. I microservizi coprono questo ruolo, hanno le stesse caratteristiche di un servizio ma eseguono solo su richiesta.

L'ambiente di esecuzione è completamente gestito da AWS, l'unica requisito per creare un microservizio è caricare il proprio codice. In questo caso viene noleggiato il tempo di calcolo invece che una macchina fisica o virtuale. L'ambiente viene creato al momento della richiesta, esegue il codice e cessa di esistere. Quando il microservizio non è attivo non ci sono costi.

È bene tenere in mente due importanti caratteristiche dei microservizi:

- L'ambiente non ha spazio di archiviazione, qualora sia necessario salvare un file, bisogna caricarlo in un servizio di cloud computing come S3
- Le tecnologie devono essere compatibili con l'infrastruttura sottostante, il che limita le nostre scelte

Avendo in mente queste considerazioni, un caso d'uso adatto ai microservizi è un servizio API che viene usato in maniera occasionale oppure per brevi periodi fissi.

2.4 Sistema automatico di generazione di report

Engim esegue una manutenzione annuale di database, che consiste nell'archiviazione dei dati. Questi possono essere salvati dal cliente, nel caso fosse interessato o necessitato, sotto forma di pdf oppure xls.

Le informazioni più importanti sono l'elenco e le specifiche di tutte le "attività". Un'attività contiene una serie di dati, tra cui: coordinate gps, costi di lavoro, tempo di lavoro e altro. Al momento il servizio è implementato da Rails tramite una gemma di ruby. Il sistema attuale crea un'istanza di chrome, l'istanza contiene un HTML che si desidera convertire in pdf e infine avviene il parsing del documento.

Questo ha una serie di gravi problemi:

- La necessità di avviare un'istanza di Chrome e il parsing di un HTML è estremamente costoso dal punto di vista delle risorse
- Il parsing di un HTML è anche estremamente costoso in termini di tempo, aggravato dalle lunghe query dovute alla grande mole di dati
- Il parsing tende a essere poco affidabile

La natura del nostro problema rende molto facile la scelta di un microservizio. L'operazione è ripetitiva, ben definita e usata per brevi periodi. Altri vantaggi importanti sono il risparmio di risorse del server, che evita di gra-

vare sulle operazioni più critiche, e la possibilità di usare il microservizio per qualsiasi altro prodotto Engim.

3 Requisiti del progetto

3.1 Descrizione

Il progetto è una funzione lambda su AWS, ovvero un microservizio. La funzione è chiamata tramite con una PUT tramite https. Il json contiene i dati di autenticazione e il contenuto da convertire in pdf. L'autenticazione avviene tramite token e whitelist, questi sono salvati come variabili di ambiente su AWS. I contenuti invece sono divisi in sezioni per permettere dinamicità di stampa, qualora uno dei blocchi fosse assente, semplicemente non verrà stampato.

Nel caso in cui la stampa sia avvenuta correttamente la funzione ritorna 200 e il link al file su S3. Se l'input della chiamata risulta errato, la funzione ritorna 400 e un messaggio che descrivere l'errore. Infine se avviene un errore di connessione al bucket S3, la funzione ritornerà un errore 500.

Il progetto deve permettere l'implementazione di stampe diverse da quelle di servizioGPS e di altri file di output come kml e xls.

3.2 Sicurezza

L'autenticazione avviene a livello di codice. AWS permette di creare variabili d'ambiente, evitando che siano scritti in chiaro nel codice. Una di queste variabili è un token di autenticazione, esso viene confrontato con il token in

input.

La whitelist invece è una lista dei domini di Engim, questi vengono caricati in una lista e confrontati con il dominio mittente.

Sarebbe stato possibile fare un'autenticazione IAM alla lambda. In questo caso bisognerebbe creare un utente per ogni server, che possa fare il login ad AWS prima di poter chiamare la funzione. Indubbiamente questa opzione è più sicura, allo stesso tempo richiederebbe la creazione e gestione di molteplici utenti, uno per ogni server presente. Inoltre solo l'amministratore ha i permessi per farlo. La poca praticità del metodo unito al fatto che la lambda non contiene alcuna informazione sensibile o in generale rischi per Engim, rende la scelta più semplice più attraente.

Qualsiasi chiamata da dominio esterni oppure senza token verrà tratta come "400 bad request".

3.3 Correttezza dei dati

Esistono 4 blocchi principali: header, body, table e printInfo. Fin tanto che almeno uno dei 3 è presente insieme al printInfo, la stampa risulta valida.

Il header può contenere fino a 2 loghi, però deve necessariamente contenere un titolo.

Il body può contenere un numero arbitrario di blocchi. Ogni blocco è racchiuso in un rettangolo e a sua volta contiene una serie di titoli con i loro dati. È valida l'assenza di un dato, ma invalida l'assenza di un titolo.

Il table contiene una serie di titoli e una serie di righe arbitrariamente lunga. Anche in questo caso non può mancare un titolo ma un dato può essere vuoto. Tra i tipi di dato è possibile avere un'immagine, l'immagine viene

spediata come base64, se è assente non verrà scritto nulla.

Il programma presume che i dati siano corretti, controlla solo la loro presenza.

3.4 Problemi interni del server

3.5 Presupposti e dipendenze

La scelta del linguaggio è limitata da AWS, in particolare i 3 presi sotto considerazione sono state ruby, python e javascript. Le librerie di ruby prevedono il parsing di HTML quindi non risolvono i problemi preesistenti. Python offre molte librerie, ma con funzionalità parziali oppure non mantenute. Javascript invece ha pdfkit, una libreria che permette la creazione e manipolazione di un PDF senza intermezzi.

Il sistema non controlla la correttezza dei dati, ma vengono controllate l'assenza dei campi necessari. Un titolo non può mancare, mentre è possibile che un campo sia vuoto.

La stampa ha una forma regolare, divisa per righe e colonne. Se i dati sono disposti in maniera più arbitraria è necessario implementarlo.

Non è previsto nel progetto altre stampe al interno del progetto corrente, ma svilupparlo in modo da permettere liberamente di aggiungere altri tipi di stampa senza difficoltà.

3.6 Features

La lambda è una funzione asincrona che prende in input un evento e ritorna a un JSON con eventuali informazioni o messaggi di errore.

4 Sviluppo

4.1 Workflow

Il mio senior ha deciso di fare un implementazione a grossi step, ogni volta facendo un piccolo pezzo per poi controllare che tutto funzioni correttamente. Ogni issue è uno step e a ogni step il progetto viene testato sia manualmente su aws tramite simulazioni di richiesta sia con test automatizzati tramite mocha.

4.2 Creazione di un file e salvataggio su S3

La lambda è una funzione asincrona che viene eseguita ogni volta che riceve una chiamata https. Bisogna prima di tutto creare una connessione con il bucket di s3, una volta creata questa connessione si può salvare un file passando il filepath oppure uno stream di data e avremo salvato il file su S3

4.3 Stampa di un file pdf da un json

Le operazioni utili al mio scopo text per scrivere una stringa, stroke per disegnare una linea e infine è possibile manipolare colori e font. Inizialmente avevo deciso di programmare in maniera funzionale perché risultava una logica più semplice essendo un breve script, con l'aumentare della complessità e con il modo di esportare di javascript ho pensato che una classe fosse più adeguata. Quindi c'è un costruttore che definisce la grandezza del documento, questo mantiene la relazione di un foglio A4 (circa 1.41) in modo che stampa e visualizzazione sono sempre molto comodi. Dopo l'inizializzazione si inizia

la scrittura dei blocchi. Ogni json potrebbe avere un header, body e table. Se uno dei 3 manca, non verrà scritto e gli altri si organizzeranno di conseguenza. Ogni metodo di scrittura controlla che non ci siano dati mancanti che possano nuocere alla stampa come ad esempio se manca un titolo non è una stampa valida, e ritornano 400. Se tutto va bene e la stampa va a buon fine, è possibile mandare l'oggetto con get document. Ogni possibile caso è testato con mocha, la connessione con AWS è mochata tramite aws-sdk e gran parte dei dati è generata casualmente tramite faker. Se tutto va a buon fine il programma ritorna 200 e il link del download.

5 Implementazione

6 Applicazione e performance

7 Conclusione

8 Appendici

9 Bibliografia