

DISTRIBUTED SYSTEMS AND BIG DATA

Relazione Homework 1- Salvatore Pulvirenti

L'obiettivo del presente lavoro è la realizzazione di un sistema con il quale degli utenti possano leggere i valori delle azioni ovvero i ticker di aziende che vengono raccolte dal sito "Capital IQ and Morningstar" e rese disponibili da YAHOO. Per rendere il sistema scalabile il progetto si fonda sulla divisione delle funzioni in microservizi che vengono inserite in un container. In particolare si è diviso il sistema in tre sottoservizi:

- Il server che si interfaccia con gli utenti
- Un database che raccoglie la tabella degli utenti, formata da id e ticker e la tabelle dei ticker richiesti.
- Un microservizio che raccoglie le informazioni dei ticker degli utenti

In figura 1 lo schema del container e della divisione dei microservizi.

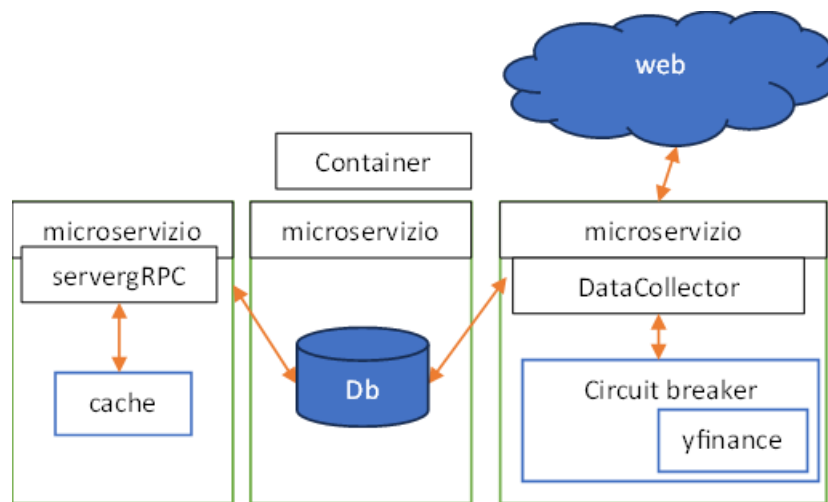


Figura 1

La scelta su cosa dovesse contenere ciascun servizio è stata fatta seguendo la funzionalità delle operazioni: servergRPC, Database per la persistenza dei dati, Datacollector per raccogliere le informazioni. Per i tre microservizi e il container è stato utilizzato **Docker compose**.

ServergRPC

La comunicazione fra il server e il client avviene tramite gRPC. Questo è un framework moderno ad alte prestazioni. Nato da Google, serve a semplificare la comunicazione fra client e sistemi di back-end. La comunicazione dal lato client sembra essere una chiamata locale ma diventa essenzialmente una chiamata out-of-process trasparente a un servizio remoto. Lo stato RPC astrae la comunicazione di rete point-to-point, la serializzazione e l'esecuzione tra computer. Per migliorare la risposta è stata aggiunta al server una cache, in cui sono tenuti gli utenti che si sono registrati, evitando che un utente possa registrarsi una seconda volta applicando la politica di “at-most-once”. Il microservizio del server espone la funzionalità di gRPC sempre alla porta 50051.

Database

La persistenza delle informazioni è stata fatta con un database MySQL inserito in un volume. Per memorizzare gli utenti è stata creata una tabella con due campi: ID che è la mail dell'utente e il ticker. Questa tabella, chiamata “USERS” viene tenuta aggiornata dal server.

Field	Type	Null	Key	Default	Extra
email	varchar(255)	NO	PRI	NULL	
ticker	varchar(10)	NO		NULL	

La seconda tabella è quella dei ticker chiamata in seno al progetto “stock_data”, viene aggiornata dal datacollector con i dati richiesti.

Field	Type	Null	Key	Default	Extra
Id	Int	NO	PRI	NULL	Auto_increment
email	varchar(255)	YES	MUL	NULL	
ticker	varchar(10)	YES		NULL	
Value	Double	NO		NULL	
timestamp	Datetime	NO		NULL	

Datacollector

A questo servizio è stato demandato il compito di raccogliere i dati e di rendere affidabile la comunicazione con il sito web da dove vengono presi i dati. Per fare questo è stato inserito nel codice del datacollector un algoritmo denominato “circuit-

breaker". Il nome deriva dagli interruttori magnetotermici di rete, che si staccano in presenza di un problema e poi essere riarmati poco dopo. Infatti quello che fa questo algoritmo è esattamente questo che ci prova a comunicare con il sistema remoto e in mancanza di risposta si apre per riprovarci poco dopo. I circuit-breaker vengono inseriti fra l'applicazione e il sistema remoto e possono essere implementati come una macchina a stati e tre stati sono:

- Chiuso: se in questo stato si verificano un certo numero di errori per unità di tempo il circuit breaker passa allo stato di aperto. Trascorso un certo tempo di timeout il circuit breaker passa allo stato di semi aperto.
- Aperto: la richiesta non è andata a buon fine e viene restituito all'applicazione un errore.
- Semiaperto: solo un numero limitato di richieste dell'applicazione può passare. Se queste hanno esito positivo il circuit breaker passa allo stato di chiuso, se invece una richiesta non va a buon fine passa allo stato di aperto.

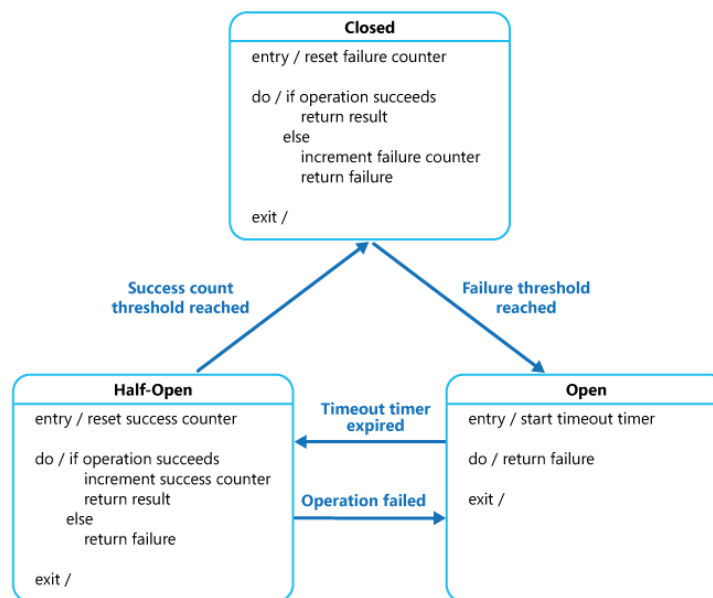


figura 2

Per raccogliere i dati richiesti è stata usata la libreria yfinance.

Test del sistema.

Per testare le funzionalità del sistema è stato scritto un client che consente di registrare gli utenti e prelevare i dati dei ticker. Benchè il sistema funzioni sono state rilevati dei malfunzionamenti che andranno rivisti e corretti.