# 

# **ALMA MATER STUDIORUM - UNIVERSITA' DI BOLOGNA CAMPUS DI CESENA**

DIPARTIMENTO DI INFORMATICA – SCIENZA E INGEGNERIA

CORSO DI LAUREA IN INGEGNERIA E SCIENZE INFORMATICHE

**TITOLO DELL’ELABORATO**

*Elaborato in*

BASI DI DATI AVANZATE

*Relatore* *Presentata da*

**Prof. Alessandra Lumini** **Francesco Foschini**

Anno Accademico 2018 - 2020

# Sommario

Lo studio eseguito prevede la realizzazione di un sistema che permetta di distribuire notifiche che si riferiscono a eventi generati sui software di backend e di frontend aziendali, per fornire informazioni costantemente aggiornate agli utenti delle applicazioni web e monitorare lo stato di avanzamento di task/procedure.

Per il design del sistema è stato necessario acquisire una solida conoscenza dei sistemi distribuiti, dei problemi di concorrenza sui dati condivisi e dell’interfacciamento con database real-time.

Il progetto ha previsto l’implementazione del server e della libreria di sistema UtilityRethink. In seguito ad un’attenta ricerca iniziale è stato deciso di affidare il ruolo del server di sistema al dbms RethinkDB. La libreria, invece, è stata scritta in linguaggio C# su piattaforma .NET Core allo scopo di garantire ai servizi client l’interfacciamento al server di sistema. Il codice prodotto è stato organizzato in maniera efficiente e riutilizzabile nei prodotti della società.

Dal punto di vista del database, è stato studiato lo schema migliore di rappresentazione delle notifiche per ottenere le migliori prestazioni possibili a fronte delle necessarie caratteristiche di consistenza e isolamento.

L’attività è stata svolta interamente presso l’azienda Energy Software s.r.l. di Faenza che è un’azienda nata nel 2015 grazie a un team di persone con esperienza decennale unite dall’obiettivo di fornire tutta una serie di servizi legati al mondo dell’informatica tra cui lo sviluppo di software per i mercati energetici, lo sviluppo di applicativi per specifiche esigenze del cliente e la consulenza sistematica specializzata su server e networking.

I risultati ottenuti sono stati oggetto di misurazione e presentazione.

Il seguito della tesi è articolato in 5 capitoli:

* Capitolo 1: introduce nei dettagli la collocazione di questo sistema sviluppato. Sono inizialmente trattati i nuovi sistemi dbms real-time e il design pattern Publish/Subscribe. Vengono infine definite le specifiche funzionali e architetturali del sistema.
* Capitolo 2:
* Capitolo 3:
* Capitolo 4:
* Capitolo 5:

**Indice**

[Sommario 3](#_Toc58752075)

[Specifiche e descrizione del problema 6](#_Toc58752076)

[Introduzione al problema 6](#_Toc58752077)

[Database-Realtime 6](#_Toc58752078)

[Publish/Subscribe 7](#_Toc58752079)

[Specifiche architetturali 8](#_Toc58752080)

[Specifiche funzionali 8](#_Toc58752081)

[Altri sistemi esistenti utilizzabili in futuro (Librerie e tecnologie) 9](#_Toc58752082)

[Firebase 9](#_Toc58752083)

[Database in tempo reale FireBase 10](#_Toc58752084)

[Meteor 10](#_Toc58752085)

[Parse 11](#_Toc58752086)

[RabbitMQ 13](#_Toc58752087)

[Conclusioni sulle altre tecnologie utilizzabili 14](#_Toc58752088)

[Bibliografia 15](#_Toc58752089)

**Capitolo 1**

# Specifiche e descrizione del problema

# 

## Introduzione al problema

### Database-Realtime

Gli utenti si aspettano un’immediata reattività dalle applicazioni moderne. Si presume, quindi, che qualsiasi modifica apportata nelle applicazioni si rifletta immediatamente in tutte le interfacce di ogni utente in tempo reale [[1]](#database_realtime1).

Lo sviluppo delle applicazioni reattive effettuate con la tecnologia di database tradizionale, tuttavia, presenta delle difficoltà perché i *dbms*, sistemi per l’archiviazione e il recupero dei dati, sono stati sviluppati per decenni attorno a un modello di accesso richiesta-risposta puramente “*pull-based*”.

Rispondendo alla necessità di reattività, si è sviluppata una nuova classe di sistemi *dbms* basati su un’architettura diversa dalla maggior parte dei sistemi *dbms* relazioni e non relazionali preesistenti. Invece di eseguire il *polling* per le modifiche avvenute sui dati di un database, sovraccaricando il sistema, questi nuovi *dbms* sono orientati al “*push delle modifiche”.*Lo sviluppatore, infatti, può richiedere al sistema stesso di inviare continuamente i risultati di *query* alle applicazioni client in tempo reale. Questi sistemi sono spesso definiti *database realtime* poiché mantengono i dati sul client sincronizzati con lo stato corrente del database.

L’architettura *push* in tempo reale dei sistemi *database real-time* facilita quindi lo sviluppo di applicazioni web reattive. La modifica dei dati è intercettata e notificata in tempo reale a tutti i servizi client. L’utilizzo di questi particolari e nuovi sistemi è quindi molto consigliato quando le applicazioni richiedono di essere aggiornate in tempo reale. Siccome, ad esempio, i dati correnti dei mercati della borsa di New York sono estremamente dinamici nel tempo, l’architettura offerta dai *database real-time* potrebbe essere molto vantaggiosa per un’applicazione che disegna i grafici in tempo reale di questi mercati azionari [[2]](#database_realtime2). Potrebbe inoltre essere interessante utilizzare questi sistemi nei contesti di applicazione di progettazione collaborativa, quando un utente, ad esempio, decide di cambiare la posizione di un pulsante. Sarà il *dbms* stesso a informare gli altri utenti, che stanno lavorando contemporaneamente allo stesso progetto, dell’aggiornamento effettuato. Si potrebbe infine utilizzare nello sviluppo di videogiochi multiplayer e nello streaming di applicazioni di analisi.

L’adattamento dei sistemi gestionali di database alle esigenze di aggiornamento in tempo reale rappresenta ancora un’enorme sfida ingegneristica.

Al contrario, però, i nuovi sistemi database real-time non sono sempre la scelta ottimale. Quando si ha la necessità di rispettare le proprietà *ACID* ovvero atomicità, coerenza, isolamento e durabilità dei dati occorre fare affidamento ai classici sistemi *dbms* relazionali e quando si eseguono analisi ad alta intensità di calcolo, occorre utilizzare un sistema di *Big Data* come *Hadoop* o ad un archivio orientato alle colonne come *Vertica*.

### Publish/Subscribe

Il sistema che ho sviluppato ha richiesto un unico importante e particolare requisito di partenza: il server di sistema stesso deve essere in grado di notificare eventi in real-time ai servizi client interessati. Di conseguenza, il sistema creato è un’implementazione del design pattern *Publish/Subscribe* [**[3]**](#Publish_Subscribe)**.**

Proprio per questa motivazione la tecnologia scelta per l’implementazione del server di sistema è un *dbms* real-time disponibile sul mercato: *RethinkDB*.

Nello schema *Publish/Subscribe*, il mittente dei messaggi è il server, mentre i destinatari sono i servizi client che dialogano attraverso un tramite, chiamato *dispatcher* o *broker*. Il server (*publisher*) non essendo consapevole dell'identità dei destinatari (detti *subscriber*) si limita a "pubblicare" i propri messaggi al *dispatcher*.

I client si rivolgono a loro volta al *dispatcher* abbonandosi alla ricezione di messaggi.

Il *dispatcher*, quindi, inoltra ogni messaggio inviato dal server solo ai servizi client interessati a quel messaggio specifico.

Il design offerto da *Publish/Subscribe* implica, infine, che per qualsiasi tecnologia del server scelta, non siano noti i servizi client che s’interfacciano a esso attraverso la libreria che ho implementato, contribuendo di conseguenza alla scalabilità del sistema.

Utilizzando questo pattern di messaggistica i servizi client del sistema riescono a registrarsi alla ricezione di notifiche utilizzando una funzionalità della libreria chiamata *UtilityRethink* che ho implementato (vedi capitolo *IRXNotifier*).

Il meccanismo di sottoscrizione consente, inoltre, ai servizi client di precisare a quali messaggi sono interessati: un client, ad esempio, potrebbe "abbonarsi" solo alla ricezione di notifiche aventi un determinato valore di un campo.

## 

## Specifiche architetturali

Il sistema che ho implementato coinvolge essenzialmente tre entità software principali: un applicativo client, una libreria e un server.

Il servizio client riesce a interagire con il server utilizzando la libreria.

I requisiti architetturali e tecnologici sono quindi:

• Architettura client - server;

• Accesso al server attraverso i protocolli http e https;

• Dbms non relazionale real-time

## Specifiche funzionali

La libreria di sistema implementata consente ai servizi client che ne usufruiscono di registrarsi alla ricezione di messaggi a cui si è interessati dal server.

Nel contesto aziendale, in particolare, i servizi di frontend e backend possono registrarsi alla ricezione di notifiche che si riferiscono a eventi (tabella di sistema “Notifications”) con un certo valore di argomento (campo “arg”). Se ad esempio è inserita nella tabella “Notifications”, una nuova notifica con campo “arg” uguale a “A” allora saranno avvisati della variazione tutti i servizi interessati a notifiche aventi “arg” uguale a “A”.

A seguito della scelta di una tecnologia dbms per il server di sistema è possibile, inoltre, eseguire operazioni CRUD sulle tabelle che sono considerate di sistema dalla libreria implementata *UtilityRethinnk* (es. *Notifications*).

Il sistema fornisce, infine, funzionalità di riconfigurazione delle politiche di Replication e Sharding attuate sui dati di una tabella e funzionalità gestionali per il database come la creazione e la cancellazione di nuove tabelle e di indici.

**Capitolo 2**

# Altri sistemi esistenti utilizzabili in futuro (Librerie e tecnologie)

Per l’implementazione del sistema, come spiegato nel capitolo 1, potevano essere utilizzate tutte le tecnologie in grado di implementare il design pattern *Publish/Subscribe*, come tutti gli altri sistemi dbms real-time disponibili in commercio: Firebase, Meteor, Parse. Si è utilizzato per questo progetto il dbms RethinkDB che sarà illustrato in seguito.

## Firebase

Firebase è ora la piattaforma di punta di google per la creazione di applicazioni per dispositivi mobile e web ed è stato integrato ad altri servizi google per offrire un insieme più ampio di prodotti per gli sviluppatori. Ad esempio Google cloud messaging, il servizio google per mandare notifiche push ai dispositivi Android, è superato dal prodotto FireBase Cloud Messaging, che ha aggiunto la funzionalità di invio notifiche push sia ai dispositivi iOS sia dispositivi Web [[4]](#Firebase).

Firebase Cloud Messaging (FCM) è una soluzione cloud multipiattaforma che può essere utilizzata gratuitamente; consente agli sviluppatori di applicazioni di terze parti di inviare notifiche o messaggi dai server agli utenti della piattaforma. Questo sistema consente, quindi, all’utente di ricevere messaggi di notifica o di dati che possono essere decifrati dal proprio codice dell’applicazione client. FCM può recapitare messaggi di destinazione alle applicazioni con tre metodi: a un singolo dispositivo, a un gruppo di dispositivi o ai dispositivi che sono iscritti all’argomento del messaggio [[5]](#Firebase_cloud_messaging).

### 

### Database in tempo reale FireBase

Firebase Realtime Database è un database ospitato nel cloud. I dati vengono archiviati come JSON e sincronizzati in tempo reale con tutti i client connessi.

A differenza di altri archivi di dati basati su JSON, un’istanza Firebase non è una raccolta di documenti JSON, ma un unico documento JSON: una struttura ad albero di oggetti ed elenchi nidificati. Per accedere ai dati è necessario navigare attraverso la gerarchia e richiedere specifici nodi figlio per i quali è possibile ricevere aggiornamenti immediati quando altri utenti modificano i dati.

FireBase Realtime Database permette di creare applicazioni collaborative consentendo un accesso sicuro al database direttamente dal codice lato client.

I dati sono conservati localmente e, anche offline, gli eventi in tempo reale continuano a verificarsi, offrendo all’utente finale un’esperienza reattiva. Quando il dispositivo riacquista la connessione, il database in tempo reale sincronizza le modifiche ai dati locali con gli aggiornamenti remoti che si sono verificati mentre il client era offline [[6]](#Firebase_realtime_db).

Il Real-time Database offerto da Firebase è un database NoSQL e come tale ha ottimizzazioni e funzionalità diverse rispetto a un database relazionale. L’API del database in tempo reale è progettata per consentire solo operazioni che possono essere eseguite rapidamente. Ciò consente di creare un’ottima esperienza che può servire milioni di utenti senza compromettere la reattività. Per questo motivo, è importante pensare a come gli utenti devono accedere ai dati e quindi strutturarli di conseguenza.

I dati archiviati in un Firebase Realtime Database sono recuperati collegando un listener asincrono a un riferimento del database. Il listener è attivato una volta per lo stato iniziale dei dati e di nuovo ogni volta che i dati cambiano [[7]](#Firebase_realtime_db_listener). Questa modalità di recupero dei dati è supportata in Java, Node.js e Python Admin SDK.

## 

## Meteor

Meteor è un frame work di sviluppo javascript scritto in Node.js per lo sviluppo di applicazioni moderne multipiattaforma (Web, Android, iOS).

Si integra perfettamente al dbms non relazionale MongoDB ereditando, di conseguenza, la sua espressività di interrogazione aggiungendo, inoltre, le query di aggiornamento tipiche dei sistemi database real-time [[8]](#Meteor).

Meteor utilizza il protocollo di dati distribuiti per propagare automaticamente le modifiche dei dati ai client senza richiedere allo sviluppatore di scrivere alcun codice per la sincronizzazione. Il protocollo dei dati distribuiti (DDP) è un protocollo client-server per l’interrogazione e l’aggiornamento di un database lato server e per la sincronizzazione di tali aggiornamenti tra i client [[9]](#meteor_ddp). Utilizza il modello di messaggistica Publish/Subscribe ed è stato creato per essere utilizzato dal frame work Meteor.

Il client meteor può essere implementato con qualsiasi framework JS frontend popolare: Vue, React, Angular.

## Parse

Simile a Meteor, infine, Parse è un frame work di sviluppo di applicazioni che utilizzano MongoDB come sistema dbms [[10]](#Parse1), è stato un frame work molto popolare e ha avuto una delle più grandi distribuzioni di MongoDB al mondo.

La società è stata acquistata da Facebook nel 2013 e chiusa nel gennaio 2017 divenendo open source.

Il videogioco Clash of Kings ha utilizzato Parse per l’implementazione delle notifiche push in tempo reale per gli utenti [[11]](#Parse2).

La maggior parte dei progetti che eseguono Parse provengono da applicazioni mobile (Android e iOS) e le tecnologie di frontend supportate comprendono Swift, Java, React Native, e altri [[12]](#Parse3).

Tecniche di auditing applicate ai dbms relazionali

L’unica tecnica applicata dai dbms relazionali che si avvicina concettualmente al design pattern *Publish/Subscribe* è la tecnica di auditing del database**.** Il database auditing implica, infatti, l' [osservazione di](https://en.wikipedia.org/wiki/Audit) un [database in](https://en.wikipedia.org/wiki/Database) modo da essere informati delle azioni svolte dagli [utenti](https://en.wikipedia.org/wiki/User_(computing)) [[13]](#Database_auditing).

[Gli amministratori](https://en.wikipedia.org/wiki/Database_administrators) e i consulenti di [database](https://en.wikipedia.org/wiki/Database_administrators) spesso impostano questi controlli per motivi di sicurezza al fine di vietare l’accesso ad alcune informazioni sul database a coloro che non hanno il permesso di lettura.

Attraverso le tecniche di database auditing sui dbms relazionali è quindi possibile monitorare e salvare le operazioni svolte su una tabella di un database [[14]](#Meccanismi_auditing) che diviene di conseguenza un’entità “attiva”.

Il concetto generale di database auditing riguarda il monitoraggio dell’utilizzo dei record di un database all’interno di un dbms relazionale [[15].](#SQLserver_audit)

Questa tecnica seppure si avvicini, non implementa completamente il design pattern Publish/Subscribe. Quando si controlla un database, infatti, le operazioni sui dati sono monitorate e registrate in un unico registro di controllo interno del dbms, chiamato audit trail [[16]](#why_you_need_auditTrail). Per questo motivo non è possibile scegliere di utilizzare un dbms relazionale come server di sistema perché non è garantita la notifica ai servizi client non rispettando, di conseguenza, il requisito principale del sistema da implementare.

Le tecniche di auditing hanno come obiettivo principale la sicurezza dei dati. In genere, l’auditing è utilizzato per:

* Assegnare a ogni utente delle “responsabilità” per le azioni eseguite in un particolare schema e tabella di un database.
* Impedire agli utenti del database di compiere azioni inappropriate considerando le “responsabilità” dell’utente stesso. L'implementazione degli audit trail aiuta a rendere il comportamento dell'utente più appropriato perché i suoi record utente possono essere ricondotti alla sua identità. Pertanto aiuta a prevenire fughe di dati importanti aziendali.
* Monitoraggio e raccolta di informazioni su attività specifiche del database. A volte gli audit trail possono essere utili per raccogliere informazioni statistiche.

### 

Le tecniche di auditing sono utilizzate soprattutto per prevenire la compromissione dei dati aziendali che possono essere eseguite dai dipendenti con accesso autorizzato.

L’auditing può essere soddisfatto implementando meccanismi di controllo nativi del dbms stesso oppure è possibile utilizzare software esterni dedicati. Occorre specificare, però, che l’audit nativo di ciascun dbms tende a causare un sovraccarico sul server del database; i suoi archivi di grandi dimensioni richiedono l’archiviazione del database.

I dati di audit nativi dei dbms, inoltre, non sono acquisiti nel formato richiesto dai revisori e dai team di sicurezza [[17]](#DataSunrise), pertanto, la soluzione migliore per soddisfare queste esigenze è l’integrazione di un software autonomo dedicato (es. DataSunrise, Imperva Data Activity Monitoring).

## RabbitMQ

RabbitMQ è sicuramente una diversa tecnologia che potrebbe essere utilizzata per implementare la gestione dello scambio dei messaggi tra le diverse entità del sistema [[18]](#RabbitMQ). Rappresenta, infatti, un [message-oriented middleware](https://it.wikipedia.org/wiki/Message-oriented_middleware" \o "Message-oriented middleware)  chiamato [broker di messaggistica](https://it.wikipedia.org/w/index.php?title=Broker_di_messaggistica&action=edit&redlink=1) che andrebbe a svolgere il ruolo di server del sistema implementato.

Un [broker di messaggi](https://en.wikipedia.org/wiki/Message_broker) è un programma intermedio che traduce un messaggio dal protocollo di messaggistica assunto dal mittente al protocollo di messaggistica del ricevitore. Rappresenta, inoltre, un modello architetturale per il routing dei messaggi implementando il pattern Publish/Subscribe. Il suo scopo principale è di recepire i messaggi in arrivo dalle applicazioni publisher ed eseguire alcune azioni su di essi. Ad esempio, il broker di messaggi può essere utilizzato per gestirne una coda da più ricevitori, fornendo memoria affidabile, garantendone, soprattutto, la consegna.

Nel tipico scenario dello scambio di messaggi tra le diverse entità coinvolte, RabbitMQ introduce, oltre la presenza del Publisher, del Consumer e della Queue, un nuovo elemento chiamato Exchange. Il Publisher definito da RabbitMQ non invia più il messaggio direttamente alla coda, ma passa dall’Exchange, il quale crea a sua volta la comunicazione con la coda stessa.

RabbitMQ offre la possibilità di utilizzare diversi tipi di Exchange in base al proprio contesto.

Si possono, tuttavia, mettere in evidenza tre principali categorie:

* Fanout
* Direct
* Topic

Ognuna di queste categorie definisce un diverso comportamento rispetto a come è indirizzato il messaggio dall’Exchange alle code. Utilizzando l’Exchange viene a determinarsi il sistema definito dal pattern Publish/Subscribe. Nella realtà, spesso, non sarà presente un solo consumatore ma migliaia, per questo motivo il messaggio pubblicato dal Publisher è inviato a tutte le code sottoscritte a uno specifico Exchange relativo al Publisher mittente.

## 

## 

## Conclusioni sulle altre tecnologie utilizzabili

In conclusione, le tecnologie utilizzabili per l’implementazione del server di sistema sono tutti i software e applicativi in grado di implementare il design pattern *Publish/Subscribe*: database real-time o RabbitMQ. Nel mio progetto ho scelto di sviluppare il design del sistema indipendentemente dal tipo di tecnologia scelto per il server.

Se in futuro fosse, quindi, necessario utilizzare un diverso sistema dbms o un’altra tecnologia si potranno applicare poche e semplici modifiche di refactoring sulle classi di libreria implementate di UtilityRethink (vedi capitolo successivo).

**Capitolo 6**

# Bibliografia

1. Database real-time. URL <https://medium.baqend.com/real-time-databases-explained-why-meteor-rethinkdb-parse-and-firebase-dont-scale-822ff87d2f87>.
2. Database real-time. URL <https://en.wikipedia.org/wiki/Real-time_database>.
3. Design Publish/Subscribe. URL <https://it.wikipedia.org/wiki/Publish/subscribe>.
4. Firebase. URL <https://it.wikipedia.org/wiki/Firebase>
5. Firebase Cloud Messaging. URL <https://en.wikipedia.org/wiki/Firebase_Cloud_Messaging#Web_Push_Support>
6. Firebase Real-time Database. URL <https://firebase.google.com/docs/database>
7. Firebase Real-time Database listener. URL <https://firebase.google.com/docs/database/admin/retrieve-data>
8. Meteor. URL <https://en.wikipedia.org/wiki/Meteor_(web_framework)>
9. Protocollo Meteor DDP. URL <https://github.com/meteor/meteor/blob/devel/packages/ddp/DDP.md>
10. Parse. URL <https://en.wikipedia.org/wiki/Parse_(platform)>
11. Parse. URL <https://medium.baqend.com/parse-is-gone-a-few-secrets-about-their-infrastructure-91b3ab2fcf71>
12. Parse. URL <https://www.back4app.com/parse>
13. Database auditing. URL <https://en.wikipedia.org/wiki/Database_audit>.
14. Meccanismi di auditing. URL [https://www.ugiss.org/2008/09/17/meccanismi-di- auditing/](https://www.ugiss.org/2008/09/17/meccanismi-di-   auditing/).
15. SQL Server Audit (Database Engine). URL <https://docs.microsoft.com/it-it/sql/relational-databases/security/auditing/sql-server-audit-database-engine?view=sql-server-ver15>.
16. Why you need a Database audit trail. URL <https://www.imperva.com/blog/why-you-need-a-database-audit-trail/>.
17. Esempio di software per il data-auditing. URL <https://www.datasunrise.com/blog/professional-info/why-you-need-a-database-audit-trail/>
18. RabbitMQ. URL <https://it.wikipedia.org/wiki/RabbitMQ>