

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI MATEMATICA

CORSO DI LAUREA IN
INFORMATICA

TESI DI LAUREA

Sperimentazione di Apache Kafka per l'integrazione funzionale di un'applicazione aziendale

Experimenting with Apache Kafka for the Integration of an Enterprise
Application

Relatore:

PROF. TULLIO VARDANEGA

Laureando:

ANDREA DORIGO
1170610

Anno Accademico 2020/2021

Indice

1	Contesto aziendale	3
1.1	Azienda ospitante	3
1.2	Dominio applicativo	5
1.3	Processi interni e strumenti organizzativi	6
1.4	Ambiente di lavoro	8
2	Apache Kafka nel settore dell'Integrazione Aziendale	10
2.1	L'evoluzione delle architetture di integrazione	10
2.2	Apache Kafka come Middleware	10
2.3	Il percorso di Stage	12
2.4	Motivazioni e obiettivi personali	13
3	Apache Kafka in caso d'uso simulato	14
3.1	Pianificazione del percorso e <i>Way of working</i>	14
3.2	Formazione e sviluppo del prodotto	14
3.3	Risultati raggiunti dal prodotto	14
4	Obiettivi soddisfatti e bilancio formativo	15
4.1	Obiettivi soddisfatti dallo stage	15
4.2	Maturazione professionale acquisita	15
4.3	Distanza tra le competenze necessarie e quelle acquisite nel corso di studi	15

Elenco delle figure

1.1	Logo dell'azienda ospitante lo <i>stage</i>	3
1.2	certificazioni ISO LL-C di Sync Lab	3
1.3	Concetti principali del EAI_g	5
1.4	<i>Kanban Board</i> del progetto di <i>stage</i>	7
1.5	Esempio di un'attività del processo di Formazione	8
2.1	Illustrazione di un sistema a servizi con Apache Kafka	11

Capitolo 1

Contesto aziendale

1.1 Azienda ospitante



Figura 1.1: Logo dell'azienda ospitante lo *stage*

Fonte: <https://www.synclab.it/>

Sync Lab s.r.l.¹ è un'azienda di produzione software, ICT_g² e consulenze informatiche nata nel 2002 a Napoli. L'azienda al suo stato attuale presenta un organico aziendale composto da più di 200 risorse, con un fatturato annuo di 12 milioni, una solida base finanziaria e una diffusione sul territorio a livello nazionale attraverso cinque sedi: Napoli, Roma, Milano, Padova e Verona. Sync Lab possiede delle significative fette di mercato riguardanti lo sviluppo di prodotti nel settore mobile, videosorveglianza e sicurezza delle strutture informatiche aziendali.

L'azienda ha acquisito numerose certificazioni ISO LL-C per attestare la qualità dei servizi forniti, come ISO-9001 per la gestione della qualità, ISO-14001 per la gestione dell'ambiente, ISO-27001 per la sicurezza dei sistemi di gestione dati e ISO-45001 per la sicurezza nel luogo di lavoro.



Figura 1.2: certificazioni ISO LL-C di Sync Lab

Fonte: <https://www.synclab.it/>

¹Sync Lab. URL: <https://www.synclab.it/>.

²Information and Communication Technologies

Sync Lab offre i seguenti servizi per i clienti:

- Valutazione e controllo progetti
 - *Planning e project management*; definizione di *Milestone* e *team* di progetto.
 - Valutazione di impatto e *risk analysis*; monitoraggio e *benchmarking*.
 - Valutazione e controllo di progetti *software* attraverso l'utilizzo di metriche e modelli economici di stima e previsione.
- Sistemi distribuiti di *Enterprise*
 - Progettazione e realizzazione di sistemi distribuiti *Enterprise* in architettura J2EE³, EJB_g⁴, COBRA_g⁵ e *Web Services*.
 - Progettazione e realizzazione di sistemi basati su MOM_g⁶ e JMS_g⁷.
- Tecnologie *Object Oriented*
 - Applicazione delle tecnologie O-O_g all'analisi e progettazione di *software* applicativo e di sistema e nella definizione di architetture distribuite *enterprise*.
 - Utilizzo di metodologie O-O_g per progettazione di applicazioni e processi e UML, con supporto di strumenti di *modeling*, applicazione e definizione di *Design Pattern*.

L'azienda è suddivisa in molteplici settori dislocati nelle diverse sedi; l'esperienza personale mi ha portato a conoscere il settore dell'*Enterprise Architecture Integration* e del *Technical Professional Services Padova*.

Tra i clienti di Sync Lab vi sono ditte a livello nazionale di grandi dimensioni e ampio organico, come Intesa San Paolo, TIM, Vodafone, Enel e Trenitalia che necessitano prodotti sicuri e adatti al considerevole flusso di dati aziendale. L'azienda ha fornito prodotti e consulenze a più di 150 clienti, distribuiti tra clienti diretti e finali.

³Java 2 Platform Enterprise Edition

⁴Enterprise Java Bean

⁵Common Object Request Broker Architecture

⁶Message Oriented Middleware

⁷Java Message Service

1.2 Dominio applicativo



Figura 1.3: Concetti principali del EAI_g

Fonte: <https://commons.wikimedia.org/wiki/File:KrisangelChap2-EAI.png>

Il percorso di *stage* intrapreso è associato al settore del *Enterprise Architecture Integration*, che si occupa principalmente dell'EAI_g (*Enterprise Application Integration*) ovvero dell'integrazione funzionale di applicazioni aziendali per una clientela di grandi dimensioni (come un'azienda di telecomunicazioni), tramite sistemi di integrazione e *Middleware*.

I *Middleware* e sistemi di integrazione prodotti comprendono l'utilizzo di molteplici linguaggi e tecnologie in continua evoluzione.

Dal sito di Red Hat⁸:

Il middleware è un software che fornisce alle applicazioni servizi e capacità frequentemente utilizzati, tra cui gestione dei dati e delle API, servizi per le applicazioni, messaggistica e autenticazione.

Aiuta gli sviluppatori a creare le applicazioni in modo più efficiente e agisce come un tessuto connettivo tra applicazioni, dati e utenti.

Può rendere conveniente lo sviluppo, l'esecuzione e la scalabilità di applicazioni alle organizzazioni con ambienti multicloud e containerizzati.

⁸ Cos'è il Middleware? URL: <https://www.redhat.com/it/topics/middleware/what-is-middleware>.

I *Middleware* pertanto vedono due importanti utilizzi nel settore EAI_g:

- **Integrazione su più livelli:** i *Middleware* connettono i principali sistemi aziendali interni ed esterni. Capacità di integrazione quali trasformazione, connettività, componibilità e messaggistica enterprise, abbinate all'autenticazione SSO, aiutano gli sviluppatori a estendere tali capacità su diverse applicazioni.
- **Flussi di dati:** le API rappresentano una modalità per condividere i dati tra le applicazioni. Un altro approccio è quello del flusso di dati asincrono, che consiste nella replica di un set di dati in un livello intermedio, da cui i dati possono essere condivisi con più applicazioni.
- **Ottimizzazione di applicazioni esistenti:** con l'adozione del *Middleware*, gli sviluppatori possono trasformare le applicazioni monolitiche esistenti in applicazioni *cloud native*, mantenendo i validi strumenti già in uso ma migliorandone prestazioni e portabilità.

Questo contesto dell'integrazione aziendale porta dunque l'impresa ad avere un'importante propensione all'innovazione, talvolta esplicitamente richiesta dai clienti. Una direzione dell'evoluzione attuale nel settore EAI_g riguarda la migrazione verso sistemi sempre più distribuiti, in grado di gestire efficacemente ed in tempo reale flussi di dati in continua crescita. L'avanguardia tecnologica è uno dei principali temi dell'azienda, che garantisce che essa rimanga sempre competitiva sul mercato dei sistemi di integrazione (nel settore dell'EAI_g).

1.3 Processi interni e strumenti organizzativi

L'azienda adotta dei processi interni per delineare l'avanzamento di un progetto. Sync Lab utilizza una strategia *Agile* per gestire i progetti in modo da consentire un'evoluzione e adattabilità in base alle richieste del cliente e fornire soluzioni *ad-hoc*. Durante il percorso di *stage* sono stato coinvolto nei processi di Formazione, Sviluppo, Verifica e Collaudo; i processi di Manutenzione ed Evoluzione sono stati solamente accennati in quanto al di fuori dello scopo del percorso. Questi processi, nella mia esperienza personale, non sono stati delineati rigorosamente: ciò per garantire libertà e flessibilità allo stagista e all'intero progetto. Ogni processo è suddiviso in attività modulari, per rendere l'avanzamento efficace e quantificabile.

L'organizzazione efficiente di un progetto è garantita dall'utilizzo dei vari strumenti a supporto, quali *Kanban Board* (come *Click Up*⁹ per la gestione di progetto e *Notion*¹⁰ per le prenotazioni

⁹ *ClickUp™* | One app to replace them all. URL: <https://clickup.com>.

¹⁰ *Notion* - The all-in-one workspace for your notes, tasks, wikis, and databases. URL: <https://www.notion.os>.

della postazione di lavoro in sede), *chat* (come *Google Chat*¹¹) per i confronti rapidi con gli altri membri interni al progetto ed e-mail per le comunicazioni con componenti esterni al progetto.

Lo strumento più utilizzato in ambito organizzativo durante il mio percorso è la *Kanban Board* di *Click Up*, che ha permesso la gestione, il confronto, la quantificazione e la verifica del progresso. La figura seguente illustra uno *screenshot* che raffigura lo stato dell'avanzamento.

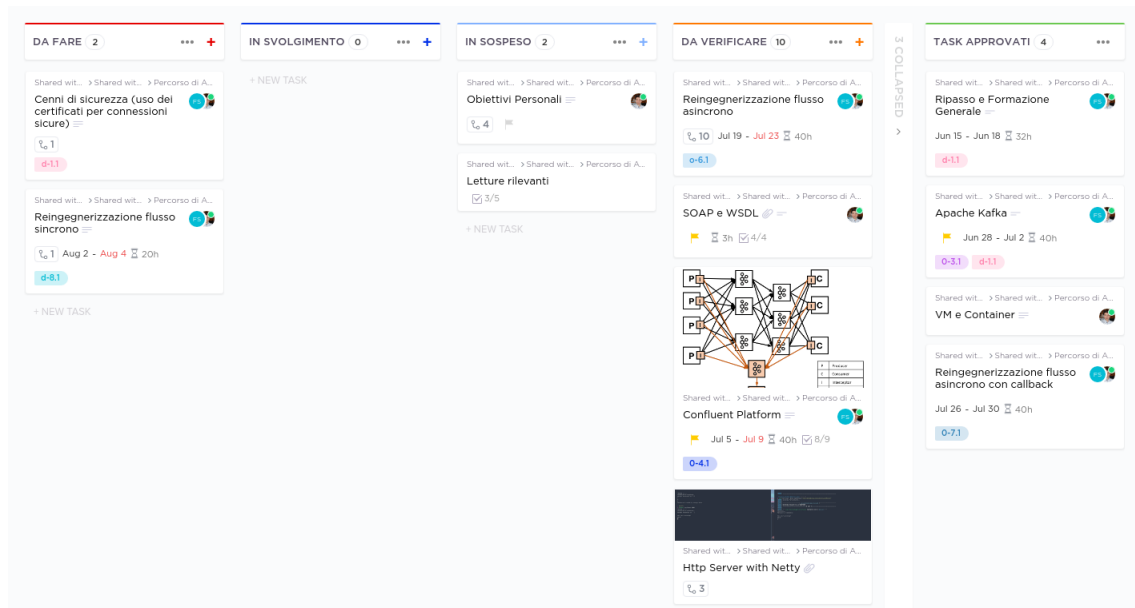


Figura 1.4: *Kanban Board* del progetto di *stage*

Le attività (*task*) vengono inizialmente create nella colonna "DA FARE" dal tutor aziendale o da me, ove ritenuto opportuno. Per dimostrare l'avanzamento il *task* si sposta verso destra a seconda dello stato raggiunto; lo stagista ha la responsabilità del cambiamento di stato fino alla colonna "DA VERIFICARE", dopodichè è compito del tutor aziendale la verifica e lo spostamento del *task* in "TASK APPROVATI", che comporta l'approvazione finale e conclusione dell'attività.

Per tenere traccia del lavoro svolto riguardante una specifica attività ho utilizzato le *card* messe a disposizione dalla piattaforma, che mi hanno consentito di delineare precisamente la pianificazione e descrizione dell'avanzamento in dettaglio del singolo *task*.

¹¹ *Google Chat*. URL: <https://chat.google.com>.

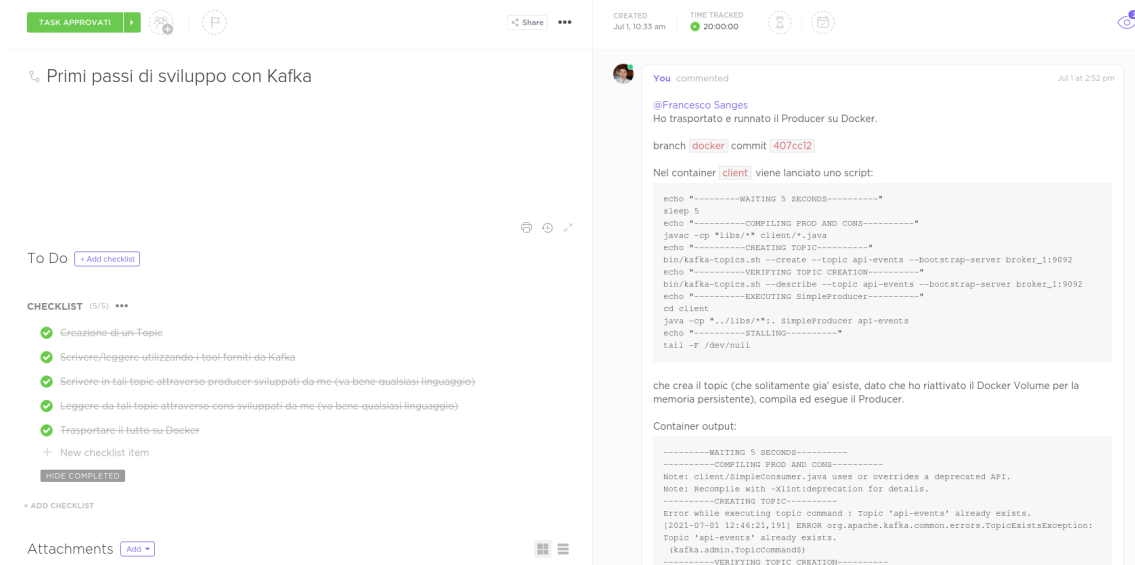


Figura 1.5: Esempio di un'attività del processo di Formazione

Questa *card* contiene una casella di testo per inserire una descrizione e appunti utili ove sia richiesto, una *checklist* approfondita, e una colonna che mantiene uno storico dei commenti; quest'ultima colonna non solo permette a me di mantenere un'importante resoconto sul lavoro svolto, ma consente anche al tutor aziendale e esperti del settore di quantificare il progresso e di fornire un aiuto rapido e contestuale.

1.4 Ambiente di lavoro

L'ambiente di lavoro di cui ho avuto esperienza risulta libero e flessibile. Lo sviluppo del prodotto nell'ambito del EAI_g dev'essere indipendente dal linguaggio di programmazione, dagli strumenti utilizzati per l'esecuzione e sviluppo, e possibilmente anche dal Sistema Operativo su cui eseguire il *software*. A tal scopo si utilizzano strumenti quali *Virtual Machine* e *Container*: essi non solo garantiscono l'indipendenza dal Sistema Operativo in uso, ma simulano efficacemente il caso d'uso reale in cui gli eseguibili sono dislocati in più dispositivi come spesso accade per il cliente finale.

Nonostante il percorso formativo abbia visto l'apprendimento di entrambe le tecnologie tramite l'utilizzo dei *software Virtual Box*¹² e *Docker*¹³, solo quest'ultima è stata utilizzata durante il progetto poiché più efficiente e minimale. Più precisamente, ho utilizzato l'estensione *Docker-compose*¹⁴ per gestire in modo elegante la generazione e collaudo di più servizi indipendenti: non

¹² Oracle VM Virtual Box. URL: <https://www.virtualbox.org/>.

¹³ Empowering App Development for Developers | Docker. URL: <https://www.docker.com/>.

¹⁴ Overview of Docker Compose | Docker Documentation. URL: <https://docs.docker.com/compose/>.

solo questo *software* consente di creare una rete di *container* comunicanti, ma rende anche rapido ed efficiente lo sviluppo grazie alla possibilità di modificare e riavviare un singolo servizio all'interno del sistema.

Capitolo 2

Apache Kafka nel settore dell'Integrazione Aziendale

2.1 L'evoluzione delle architetture di integrazione

Per soddisfare le richieste clientelari ed essere sempre competitiva e all'avanguardia, una priorità aziendale sono le esplorazioni tecnologiche e di prodotto anche tramite l'utilizzo di percorsi di *stage* insieme ai laureandi, come quanto accaduto nella mia esperienza.

Nel settore dell'*Enterprise Application Integration* in particolare, l'evoluzione tecnologica è diretta verso soluzioni sempre più distribuite e con un flusso di dati in continuo aumento. Uno degli obiettivi specifici dell'area EAI di Sync Lab è pertanto quello di trovare un software o tecnologia in grado di soddisfare i bisogni dei clienti di gestire un flusso di dati di dimensioni molto maggiori a quelle attuali, tramite architetture a messaggio che utilizzano servizi distribuiti.

2.2 Apache Kafka come Middleware

Per soddisfare le esigenze di innovazione l'azienda ha avviato un percorso per indagare le capacità del software Apache Kafka nell'ambito dell'integrazione aziendale.

Kafka è una piattaforma di *event streaming*, un sistema moderno e distribuito basato sugli eventi anziché su di una soluzione più classica come può essere quella del *request/response*. Apache Kafka si integra ottimamente in molti sistemi basati sulle Architetture a messaggio, in cui lo scambio affidabile di dati in tempo reale è la priorità.

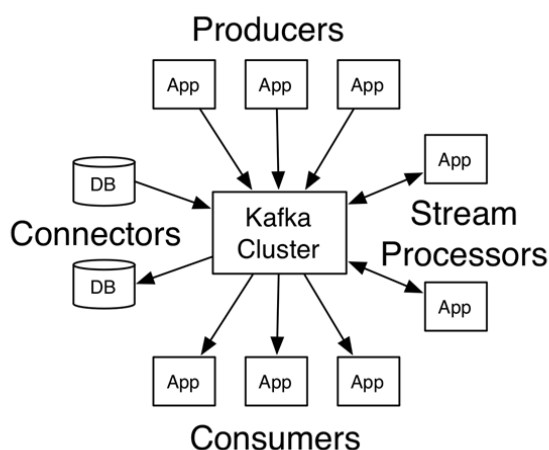


Figura 2.1: Illustrazione di un sistema a servizi con Apache Kafka

Fonte: <https://kafka.apache.org/20/documentation.html>

Il software ha dimostrato negli anni recenti un notevole successo in diversi campi, come quello del flusso di *Big Data*, del monitoraggio e dell'elaborazione dati in tempo reale. L'adozione del *software* nell'ambito dell'EAI è in crescita dato le dimostrate qualità nel gestire grandi moli di dati: la sua performance, sicurezza e scalabilità sono i punti che hanno portato il software al suo attuale successo.

L'interesse di Sync Lab nel *software* risiede dunque nell'utilizzo di Kafka come un *Middleware* per soddisfare i problemi di integrazione aziendale e reingegnerizzare i flussi di dati preesistenti, ovvero sviluppare un nuovo sistema che consenta la comunicazione tra differenti servizi con un rapido flusso di dati fra di essi. L'azienda ha avviato un percorso per testare le capacità di Kafka rispetto agli attuali strumenti utilizzati nel settore, per valutare i vantaggi e svantaggi che l'adozione di tale software può fornire al cliente.

Sono numerosi i vantaggi che Kafka può portare nel settore, fra cui:

- gestione rapida e performante di un enorme flusso di dati;
- scalabilità;
- sicurezza riguardo la persistenza dei dati;
- semplice integrazione e affiancamento a sistemi già esistenti;
- l'essere una piattaforma *open source*;
- processazione dei dati in tempo reale integrata.

2.3 Il percorso di Stage

Il percorso di *stage* offerto dall'azienda si inserisce all'interno della strategia aziendale più ampia descritta qui sopra. Al fine di esplorare la tecnologia di Apache Kafka nell'ambito di un *Middleware* per l'integrazione aziendale, l'azienda ha proposto un percorso di stage il cui obiettivo è la reingegnerizzazione di un flusso di dati asincrono, utilizzando un'architettura basata su Kafka all'interno di un caso d'uso simulato tramite servizi indipendenti. Lo stagista ha il compito di osservare, testare e verificare che il *software* possa svolgere alcuni compiti inerenti all'area dell'EAI, analizzando alcuni casi d'uso presenti in un *Middleware* aziendale in ambito *telco*.

Il percorso proposto prevede le seguenti attività e obiettivi generali:

- formazione riguardo la piattaforma di *event streaming* Kafka e l'ambito dell'integrazione aziendale;
- verifica delle capacità di Kafka nell'EAI;
- sperimentazione e sviluppo di un *Middleware* basato su Kafka tramite una simulazione di un caso d'uso reale a servizi indipendenti.

I prodotti attesi al termine dello *stage* sono dunque associati alla realizzazione di tre flussi di integrazione, basati su dei casi d'uso reali, per la gestione dei paradigmi di integrazione asincrono e asincrono con Callback (due requisiti obbligatori), e sincrone ove fosse disponibile del tempo aggiuntivo e se ritenuto opportuno; durante il percorso considerato il contesto e le opportunità offerte dal *software* questo obiettivo verrà sostituito per testare delle funzionalità aggiuntive di Kafka.

La realizzazione di questi prodotti necessita una sostanziale formazione dello stagista riguardo i principali concetti del settore del *Enterprise Application Integration* e l'utilizzo della piattaforma di *event streaming* Kafka. Più precisamente, i contenuti formativi previsti durante questo percorso di *stage* sono i seguenti:

- Concetti chiave di Enterprise Application Integration
- Design architetturali
- Cenni di Networking applicato alle architetture distribuite
- Architetture di Integrazione e Middleware
- Apache Kafka

2.4 Motivazioni e obiettivi personali

La ragione principale che mi ha portato a scegliere questo percorso di *stage* è l'interesse verso Apache Kafka, la cui formazione ritengo possa arricchire fortemente le mie capacità professionali. L'utilizzo della piattaforma di *event streaming* è sempre più in crescita, come l'evoluzione verso sistemi sempre più distribuiti e a microservizi. Sono pertanto interessato a sviluppare la mia formazione riguardo l'utilizzo e le implicazioni di questa tecnologia in rapida espansione, la cui formazione potrà essermi utile in molti campi anche al di fuori degli obiettivi dell'azienda ospitante lo *stage*. Altri fattori fondamentali alla scelta del percorso sono stati la familiarità con l'azienda, il personale giudizio positivo che ho avuto riguardo il loro metodo di lavoro e la libertà di sviluppo concessa: ho ritenuto importante la possibilità di elaborare personalmente un'architettura del caso d'uso con una visione ad alto livello, anziché il semplice sviluppo di un software predeterminato e dal percorso strettamente imposto.

Capitolo 3

Apache Kafka in caso d'uso simulato

3.1 Pianificazione del percorso e *Way of working*

Descrizione del Piano di Lavoro presentato, le strategie e il metodo di lavoro stabilito.

3.2 Formazione e sviluppo del prodotto

Descrizione delle fasi del percorso, con le sfide e problemi che esso ha presentato.

3.3 Risultati raggiunti dal prodotto

Confronto con i requisiti posti nel piano di lavoro, valutazione riguardo i risultati raggiunti, quelli non raggiunti e i possibili sviluppi futuri

Capitolo 4

Obiettivi soddisfatti e bilancio formativo

4.1 Obiettivi soddisfatti dallo stage

Valutazione oggettiva riguardo il percorso e i risultati raggiunti da esso.

4.2 Maturazione professionale acquisita

Descrizione delle conoscenze e abilità professionali acquisite grazie al percorso di *stage*. Valutazione del miglioramento personale portato avanti durante il percorso di stage.

4.3 Distanza tra le competenze necessarie e quelle acquisite nel corso di studi

Breve valutazione riguardo le difficoltà riscontrate, e considerazioni riguardo le competenze ottenute durante il corso di laurea che più mi hanno aiutato durante il percorso.

Acronimi

COBRA Common Object Request Broker Architecture. 4

EAI Enterprise Application Integration. 2, 5, 6, 8

EJB Enterprise Java Bean. 4

ICT Information and Communication Technologies. 3

J2EE Java 2 Platform Enterprise Edition. 4

JMS Java Message Service. 4

MOM Message Oriented Middleware. 4

O-O Object Oriented. 4

Bibliografia

- [1] *ClickUp™ / One app to replace them all.* URL: <https://clickup.com>.
- [2] *Cos'è il Middleware?* URL: <https://www.redhat.com/it/topics/middleware/what-is-middleware>.
- [3] *Empowering App Development for Developers / Docker.* URL: <https://www.docker.com/>.
- [4] *Google Chat.* URL: <https://chat.google.com>.
- [5] *Notion - The all-in-one workspace for your notes, tasks, wikis, and databases.* URL: <https://www.notion.os>.
- [6] *Oracle VM Virtual Box.* URL: <https://www.virtualbox.org/>.
- [7] *Overview of Docker Compose / Docker Documentation.* URL: <https://docs.docker.com/compose/>.
- [8] *Sync Lab.* URL: <https://www.synclab.it/>.