

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	Dominio applicativo	3
1.2	L'evoluzione delle architetture di integrazione	3
1.3	Processi interni e strumenti organizzativi	5
1.4	Ambiente di lavoro	5
2	Apache Kafka nel settore dell'Integrazione Aziendale	8
2.1	Apache Kafka come Middleware	8
2.2	Il percorso di Stage	9
3	Apache Kafka in caso d'uso simulato	10
3.1	Pianificazione del percorso e <i>Way of working</i>	10
3.2	Formazione e sviluppo del prodotto	10
3.3	Risultati raggiunti dal prodotto	10
4	Obiettivi soddisfatti e bilancio formativo	11
4.1	Obiettivi soddisfatti dallo stage	11
4.2	Maturazione professionale acquisita	11
4.3	Distanza tra le competenze necessarie e quelle acquisite nel corso di studi	11

Elenco delle figure

1.1	Illustrazione di un sistema a servizi con Apache Kafka	4
1.2	<i>Kanban Board</i> del progetto di <i>stage</i>	5
1.3	Esempio di un'attività del processo di Formazione	7

Capitolo 1

Contesto aziendale

1.1 Dominio applicativo

In conclusione al percorso di studi del corso di laurea in Informatica ho effettuato lo *stage* presso *Sync Lab*. È un'azienda di produzione *software* e integrazione di sistemi che fornisce principalmente prodotti a clienti di grande dimensione, sia pubblici che privati.

L'azienda è suddivisa in molteplici settori con diverse sedi; l'esperienza personale mi ha portato a conoscere il settore dell'*Enterprise Architecture Integration* e del *Technical Professional Services Padova*. Il percorso di *stage* intrapreso è associato al primo di questi, che si occupa principalmente dell'EAI (*Enterprise Application Integration*) ovvero dell'integrazione funzionale di applicazioni aziendali per una clientela di grandi dimensioni, tramite sistemi di integrazione *Middleware*.

I *Middleware* prodotti comprendono l'utilizzo di molteplici linguaggi e tecnologie in continua evoluzione. È un contesto con un'importante propensione all'innovazione, talvolta esplicitamente richiesta dai clienti: l'evoluzione attuale riguarda la migrazione verso sistemi sempre più distribuiti, in grado di gestire efficacemente flussi di dati in continua crescita.

1.2 L'evoluzione delle architetture di integrazione

L'evoluzione del settore dell'*Enterprise Application Integration* verso soluzioni sempre più distribuite e con un flusso di dati in continuo aumento ha sviluppato nei clienti (e di conseguenza nell'azienda) un interesse verso il prodotto Apache Kafka. Il software ha dimostrato negli anni recenti un notevole successo in diversi campi; l'azienda ha interesse nel testare le capacità di Kafka nel soddisfare le esigenze dell'integrazione aziendale.

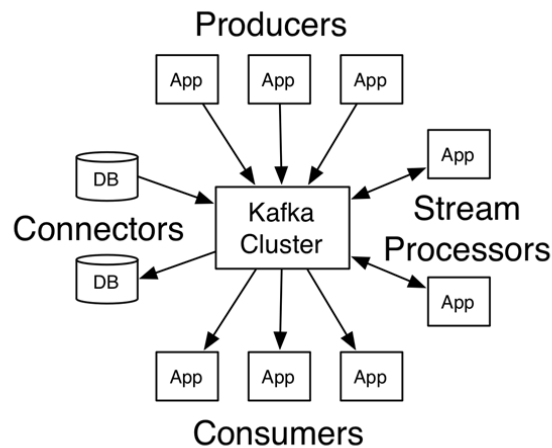


Figura 1.1: Illustrazione di un sistema a servizi con Apache Kafka

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

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*. 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 aziendale nel *software* risiede nell'utilizzo di Kafka come un *Middleware*, ovvero di un 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 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 integrata.

1.3 Processi interni e strumenti organizzativi

L'azienda adotta dei processi interni per delineare l'avanzamento di un progetto. 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, al fine di garantire una certa rapidità e adattabilità al progetto di sperimentazione. 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* per la gestione di progetto e *Notion* per le prenotazioni della postazione di lavoro in sede), *chat* (come *Google Chat*) per i confronti rapidi con gli altri membri interni al progetto ed e-mail per le comunicazioni con componenti esterni al progetto.

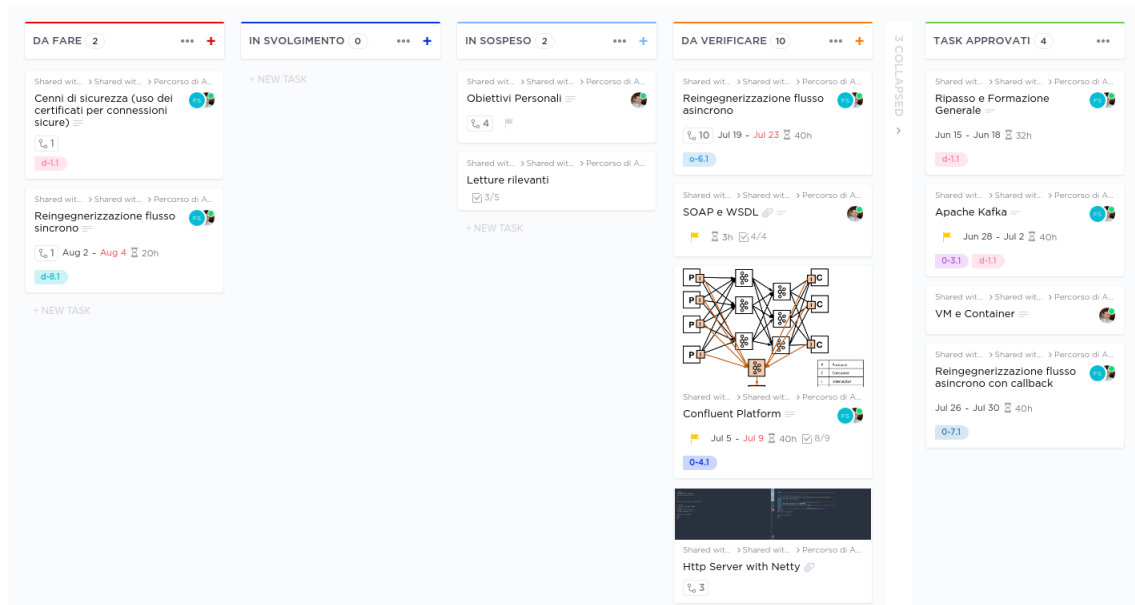


Figura 1.2: *Kanban Board* del progetto di *stage*

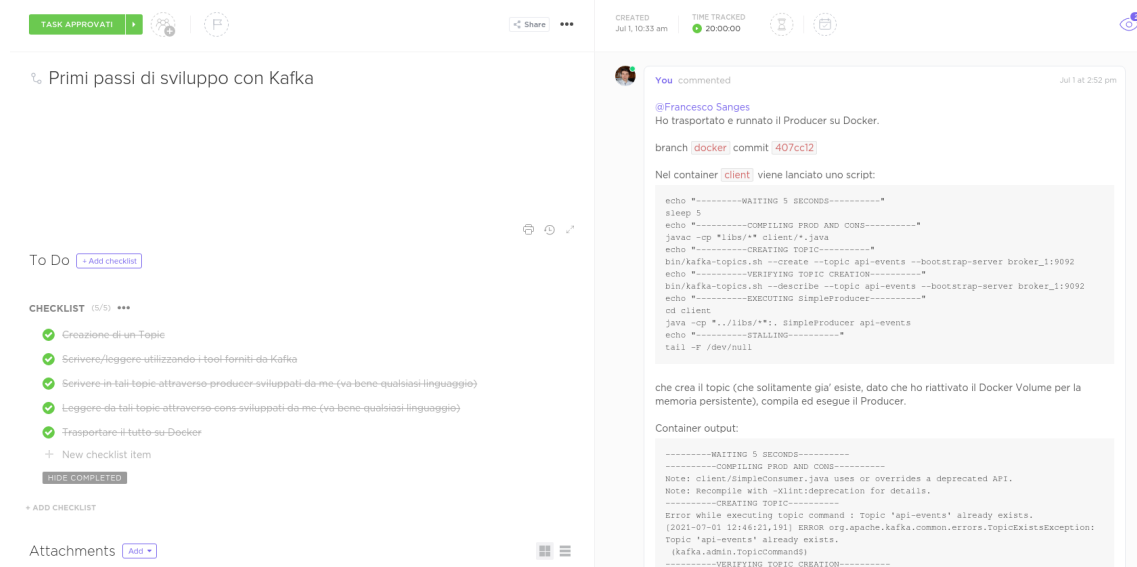


Figura 1.3: Esempio di un'attività del processo di Formazione

1.4 Ambiente di lavoro

L'ambiente di lavoro di cui ho avuto esperienza risulta abbastanza libero e flessibile. Lo sviluppo del prodotto nell'ambito del EAI 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 i vari eseguibili sono dislocati in più computer o server come spesso accade per il cliente. Nonostante il percorso formativo abbia visto l'apprendimento di entrambe le tecnologie tramite l'utilizzo dei *software Virtual Box* e *Docker*, solo quest'ultima è stata utilizzata durante il progetto poichè più efficiente e minimale.

Capitolo 2

Apache Kafka nel settore dell'Integrazione Aziendale

2.1 Apache Kafka come Middleware

L'interesse aziendale nel *software* risiede nell'utilizzo di Kafka come un *Middleware*, ovvero di un 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 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 integrata.

2.2 Il percorso di Stage

Il percorso di *stage* offerto dall'azienda si inserisce all'interno della strategia aziendale più ampia descritta qui sopra. La proposta riguarda infatti la sperimentazione di Apache Kafka in un caso d'uso simulato nell'ambito delle telecomunicazioni, ispirato ad un caso d'uso reale.

L'obiettivo principale dello *stage* è lo sviluppo di un'architettura basata su Kafka con conseguente reingegnerizzazione dei flussi di integrazioni preesistenti; lo studente 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:

- 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.

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. Altri fattori fondamentali alla scelta del percorso sono stati la familiarità con l'azienda, il loro metodo di lavoro e la libertà di sviluppo: ho valutato positivamente 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.

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.