

UNIVERSITÀ DEGLI STUDI DI MILANO  
Facoltà di Scienze e Tecnologie  
*Corso di Laurea in Informatica*

## KAFKA: STREAMING DATA

**Relatore:** Prof. Alberto MOMIGLIANO

**Correlatore:** Dr. Edmondo PORCU

Tesi di:  
Stefano CARMIGNOLA  
Matricola: 21015

Anno Accademico 2017-2018

## Indice

|  |    |
|--|----|
| <b>1. Motivazioni</b> .....  | 2  |
| <b>2. Sinossi</b> .....  | 4  |
| <b>3. ETL</b> .....  | 5  |
| <b>4. Event sourcing</b> .....                                       | 7  |
| 4.1. L'importanza dei dati e degli eventi .....                      | 7  |
| 4.2. Descrizione .....   | 8  |
| 4.3. Vantaggi .....  | 10 |
| 4.4. Svantaggi .....   | 11 |
| <b>5. Apache Kafka: Ecosistema</b> .....                             | 12 |
| 5.1. Introduzione .....  | 12 |
| 5.2. Messaggi .....  | 13 |
| 5.3. Topic e partizioni .....  | 13 |
| 5.4. Producers .....   | 14 |
| 5.5. Consumers .....   | 19 |
| 5.6. Brokers e clusters .....  | 24 |
| 5.7. Schema .....  | 26 |
| 5.8. Schema Registry .....   | 29 |
| <b>6. Architetture Event-driven</b> .....                            | 30 |
| 6.1. Kafka come piattaforma per Event Sourcing .....                 | 30 |
| 6.2. Kafka Connect: sistemi legacy, database ed Event Sourcing ..... | 33 |
| 6.3. Kafka Streams .....   | 35 |
| <b>7. Conclusioni</b> .....  | 37 |
| <b>Bibliografia</b> .....  | 41 |

## 1. Motivazioni

Nell'ultimo decennio l'avvento delle architetture a microservizi ha portato con sé la necessità di studiare nuove soluzioni al problema della gestione di molteplici fonti di dati in soluzioni software complesse.

Nei sistemi che adottano una architettura a microservizi, svariati componenti interdipendenti comunicano tra loro scambiandosi dati e attingendo da numerose sorgenti di dati comuni come *database*, *data warehouses* oppure dei servizi esterni.

La necessità di filtrare, normalizzare e gestire molte fonti di dati ha portato nei primi anni '90 alla nascita del processo di *Extract, Transform, Load* (ETL) per l'estrazione, trasformazione e caricamento di dati in sistemi di sintesi come data warehouse o data mart; questo processo si sta però rivelando complicato, impegnativo e dispendioso in un mondo dove la mole di dati prodotta dal *logging* di eventi critici per un qualsiasi business è in continua crescita: semplici esempi sono la gestione degli eventi in un sistema *Internet of Things* (IoT) oppure lo studio delle abitudini dei propri clienti per un servizio online.

Lo *stream processing* tra microservizi propone un nuovo approccio per la gestione di questi problemi, fornendo una soluzione adatta alla gestione di dati in real-time altamente scalabile e ad alto *throughput*.

Apache Kafka è una piattaforma di streaming di dati sviluppata da LinkedIn nata dalla necessità di trovare un metodo per la gestione di enormi flussi di dati (Neha Narkhede, 2017).

Kafka nasce per sfruttare a pieno lo stream processing e favorire una gestione intelligente dei dati, abbandonando il classico processo *batch* ETL per una soluzione basata sullo streaming tra microservizi favorendo un approccio improntato sulla gestione di eventi legati al comportamento dei dati, più che i dati in sé.

Durante il mio tirocinio curriculare presso Credimi<sup>1</sup> ho avuto la possibilità di partecipare allo studio ed implementazione di Kafka in vari punti del backend dell'azienda.

Credimi è una startup che opera nell'ambito fintech fornendo soluzioni di factoring per aziende. Il backend è sviluppato in Scala e richiede molteplici interazioni con sistemi esterni.

Un punto dell'architettura dove è stata studiata ed implementata una soluzione tramite Kafka è il sistema di proposta/decisione sulla validità di un utente che richiede un anticipo su una fattura caricata in piattaforma.

Al caricamento di una fattura da parte di un utente nel backend inizia un complesso processo di richieste a servizi esterni, denominati *centrali rischi*, utilizzati per reperire dati finanziari che verranno utilizzati per calcolare preventivamente l'affidabilità di questo utente e generare una *proposta di finanziamento* che sarà valutata dal back office aziendale.

In questo ambito Kafka è stato utilizzato come tecnologia alternativa ad un processo di ETL in modo da, in primo luogo, standardizzare i dati ricevuti dalle varie centrali rischi ed infine confrontare ed unire queste informazioni in un formato comprensibile ed utilizzabile dal backend per generare la proposta di finanziamento.

---

<sup>1</sup> *Credimi*, <https://www.credimi.com>

## 2. Sinossi

Prima di poter discutere della soluzione architetturale fornita da Apache Kafka e quali vantaggi propone rispetto alle soluzioni di batch ETL è necessario approfondire alcuni temi tra cui il più importante è sicuramente *event sourcing* (ES).

Il terzo capitolo presenta una veloce descrizione di un processo ETL in modo da dare visione del pattern strutturale per il quale si vuole dare una alternativa con Kafka.

Il quarto capitolo è utilizzato per illustrare uno dei concetti chiave della tesi: l'importanza di gestire e vedere le basi di dati come eventi, perché utilizzare il concetto di eventi per modellizzare i dati ed infine, come utilizzare event sourcing.

Il quinto ed il sesto capitolo sono dedicati ad esaminare la piattaforma Apache Kafka, sia da un punto di vista tecnico-architetturale esaminandone le singole parti che compongono la piattaforma, sia l'ecosistema che si è venuto a creare intorno alla piattaforma, principalmente l'utilizzo delle librerie Kafka Connect e Kafka Streams; viene inoltre presentato come utilizzare ES nel contesto dell'uso della piattaforma Kafka.

Infine nelle conclusioni vengono messi a confronto i processi ETL e la soluzione di streaming di dati offerta da Kafka nel contesto della gestione di grosse moli di dati tra microservizi e dei sistemi distribuiti.

### 3. ETL

Un processo di *Extract, Transform, Load* (ETL) è un processo mirato alla trasformazione di dati contenuti su più database per ottenere un nuovo insieme di dati, filtrato e trasformato secondo una particolare logica, destinato ad essere salvato in una data warehouse.

Verso la fine degli anni '70 molte aziende iniziarono ad utilizzare molteplici database per salvare e gestire informazioni ed è in questo contesto che nascono i processi di ETL: con l'avanzare del tempo è stato necessario studiare un metodo per l'aggregazione e gestione delle varie fonti di dati in modo tale da ottenere un insieme omogeneo di dati da utilizzare per operazioni di analisi o verifica relative al business dell'azienda.

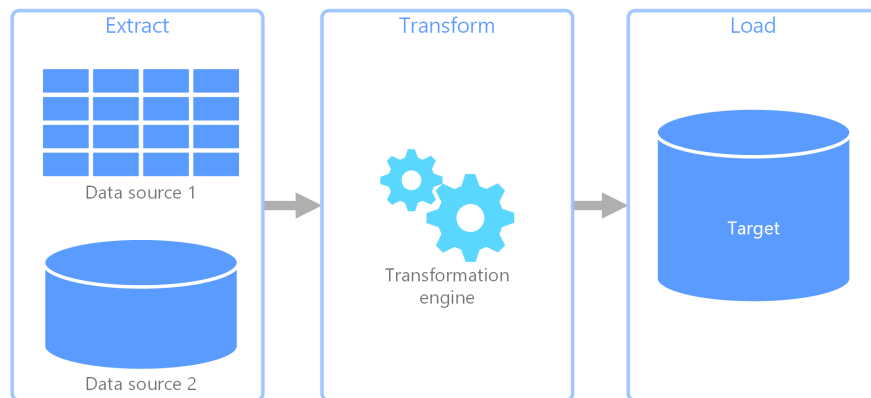


Figura 1: Rappresentazione di un processo di ETL

Un processo di ETL (Fig. 1) si compone di tre parti:

La prima parte del processo di ETL è la fase di **Extract** ed involve l'estrazione dei dati da più *data sources* come database relazionali o non-relazionali, file JSON o XML ma anche risorse *ad hoc* come, ad esempio, dei dati generati da programmi di *web analytics*.

L'obiettivo di questa fase è estrarre tutti i dati necessari dalle possibili sorgenti e prepararli alla fase di *Transform*.

Un importante problema legato a questa fase è il processo di *validazione delle sorgenti dati*: con più sorgenti dati spesso ci si ritrova a dover gestire più formati non necessariamente compatibili tra loro. Per poter garantire alla fase di Transform dei dati comprensibili, durante la fase di Extract vengono definite delle *regole di validazione* per filtrare i dati provenienti dalle varie sorgenti, un esempio di regola di validazione è il controllo dei tipi di dati presenti nella fonte.

Nella fase di **Transform** una serie di regole e funzioni vengono applicate ai dati generati dalla fase di *Extract* per prepararli alla fase di *Load* nella data warehouse.

Il primo compito della fase di *transform* è la *pulizia dei dati*: spesso le varie fonti di dati, nonostante siano state validate, possono presentare incongruenze tra loro come caratteri speciali legati all'*encoding* della propria sorgente oppure formati dei dati diversi ma compatibili (un esempio può essere la differenza di formattazione tra date americane ed europee). Per garantire un corretto funzionamento delle operazioni di trasformazione è quindi necessario pulire i dati ed adattarli ad un formato comune.

Il secondo compito della fase di *transform* è la *trasformazione dei dati* in nuovi dati richiesti dal business, esempi di trasformazioni sono:

- Joining di tabelle da più sorgenti
- Mapping e trasformazione di dati (es: "Maschio" in "M")
- Aggregazione di dati
- Generazione/calcolo di nuovi dati
- Selezione di insiemi di dati
- Validazione del nuovo formato di dati prodotto

Nella fase di **Load** l'insieme di dati generati dalla fase di *transform* vengono inseriti in un target, il quale potrebbe essere una data warehouse ma anche più semplicemente un file in un formato utile.

Business diversi hanno necessità diverse, per questo l'implementazione della fase di Load può avere più modalità implementative, il punto focale di questa fase è proprio stabilire la frequenza e le modalità di aggiornamento dei dati presenti nel target. Decidere la frequenza (giornaliera, mensile, ecc.) e le modalità (sovrascrittura dei vecchi dati o meno) del target possono portare ad un processo di ETL più o meno utile ad una azienda.

Per generare un buon target è buona norma definire uno schema *preciso e chiaro* della tipologia di dati a cui il target deve aderire.

Come detto in precedenza un processo di ETL è utilizzato per aggregare più fonti di informazioni comuni ad un processo aziendale, questo suppone che le informazioni presenti nella data warehouse potrebbero venire usate da più parti di una azienda, le quali potrebbero essere abituate a particolari formati dei dati.

Senza definire uno schema dei dati chiaro e preciso, si correrebbe il rischio di generare un insieme di dati inutilizzabile da determinati reparti in quanto non conforme al formato di dati da loro conosciuto.

## 4. Event sourcing

### 4.1. L'importanza dei dati e degli eventi

Lo status quo delle moderne applicazioni web è basato sull'utilizzo di database per rappresentare le specifiche di dominio, spesso espresse da un cliente e/o da un esperto del dominio esterno all'ambiente di sviluppo.

Durante la fase di analisi dei requisiti (supponendo un modello di sviluppo del software agile) cliente e team di sviluppo si confrontano, cercando di trovare un linguaggio comune per definire la logica e l'utilizzo del software richiesto; Una volta stabiliti i requisiti, il team di sviluppo generalmente inizia uno studio interno atto a produrre un *modello dei dati* che verrà usato come base per definire lo schema dei database utilizzati dal sistema.

Un cliente comune molto spesso non ha padronanza del concetto di 'stato di una applicazione', ma piuttosto si limita ad esporre i propri requisiti descrivendo i possibili eventi che, traslati sul modello di sviluppo incentrato su i database, portano il team di sviluppo a ragionare sui possibili stati di un database in risposta a questi eventi.

Lo stato di un database di una applicazione è strettamente legato all'insieme degli eventi del dominio applicativo; L'unico modo per modificare o interagire con questo database è tramite i comandi di inserimento, cancellazione o lettura, tutti comandi che vengono eseguiti solamente all'avvenire di un particolare evento.

Un database è spesso utilizzato solamente per mantenere lo stato corrente di una applicazione; non esiste il concetto di cronologia del database a meno di utilizzare soluzioni basate su **Change Data Capture** (CDC), generalmente utilizzate per generare un transactional log contenente tutte le operazioni eseguite sul suddetto database.

In questo modello *database-driven*, un evento genera un cambiamento su una base di dati; gli eventi e lo stato di un database sono però concetti diversi e slegati tra loro: l'esecuzione di un evento a volte può portare ad una asincronia tra l'esecuzione di un evento e lo stato di un database, tanto più se questo database è utilizzato da tutti i microservizi di una applicazione (Stopford, Messaging as the Single Source of Truth, 2017).

Una soluzione al problema di più microservizi che utilizzano lo stesso database è di utilizzare delle views del database locali ad ogni microservizio: ogni servizio lavorerà su una copia locale del database ed un job esterno si occuperà di compattare le views e mantenere il database aggiornato rispetto a tutti i cambiamenti.



Questa soluzione presenta un grosso problema: supponiamo di notare un errore sul database e di doverlo correggere, come possiamo decidere quale delle views è "più corretta" delle altre?

Per aiutarci nella ricerca dell'errore potremmo utilizzare il transactional log di ogni views, ma su database di grandezze importanti esaminare il log di ogni views che lo compone potrebbe essere un problema complesso e dispendioso in termini di tempo.

Event sourcing propone di risolvere questo genere di problemi allontanandosi da una progettazione state-driven elevando gli eventi a elementi chiavi del modello dei dati di una applicazione.

## 4.2. Descrizione

Event sourcing (ES) è un design pattern che si contrappone ad una visione del mondo basata sullo stato di una applicazione fornendo come alternativa l'uso degli eventi, ovvero delle azioni o accadimenti che l'applicazione è in grado di riconoscere e gestire come *singoli cambiamenti* dello stato del "mondo" (Stopford, Messaging as the Single Source of Truth, 2017).

La struttura dati fondamentale alla base di ES è l'**event store** una tipologia di database ottimizzata per la gestione di eventi.

In un event store gli eventi vengono inseriti in fondo alla struttura in ordine di avvenimento e non possono essere modificati o cancellati; nel caso venga pubblicato per errore un evento sbagliato o inesatto per annullarlo basterà pubblicare un evento contrario.

Questo meccanismo garantisce che la ripetizione della storia degli eventi *porterà sempre allo stesso stato, errore compreso*.

Un event store è comunemente implementato utilizzando un **log** (*Fig. 2*), una sequenza di record append-only e totalmente ordinata in base al tempo di scrittura del record (Stopford, Using Apache Kafka as a Scalable, Event-Driven Backbone for Service Architectures, 2017).

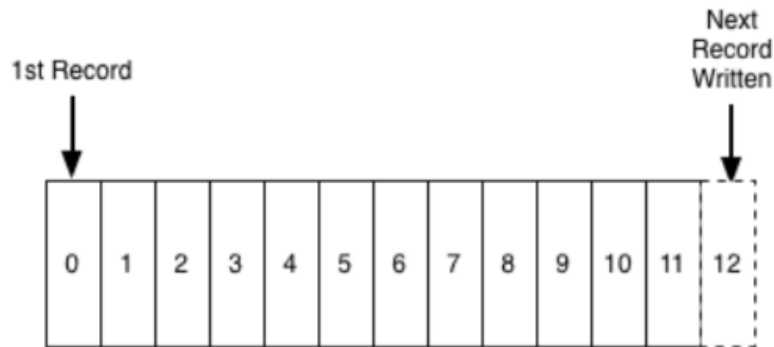


Figura 2: Struttura di un log

I record sono inseriti in fondo al log e il processo di lettura è eseguito partendo dall'inizio del log.

Generalmente in un processo di sviluppo basato su ES, si tende a nominare gli eventi con il tempo passato per esplicitare il concetto che un evento è un avvenimento passato, un esempio nome per un evento potrebbe essere **item\_created** oppure **item\_bought**.

L'ordine di pubblicazione degli eventi è di estrema importanza in quanto è ciò che permette al pattern di rappresentare correttamente lo stato di una applicazione.

È possibile vedere lo stato corrente di una applicazione come una *sequenza di operazioni di modifica dello stato eseguite partendo da uno stato iniziale*, questo implica che è possibile vedere un evento come il *delta tra lo stato iniziale di una applicazione e lo stato corrente dell'applicazione dopo l'esecuzione dell'evento* (Fig. 3).

La possibilità di trasformare lo stato corrente di una applicazione in una funzione dello stato iniziale dell'applicazione e una sequenza di eventi è il meccanismo che permette ad event sourcing di avere una validità tecnica per la gestione dei dati di una applicazione.

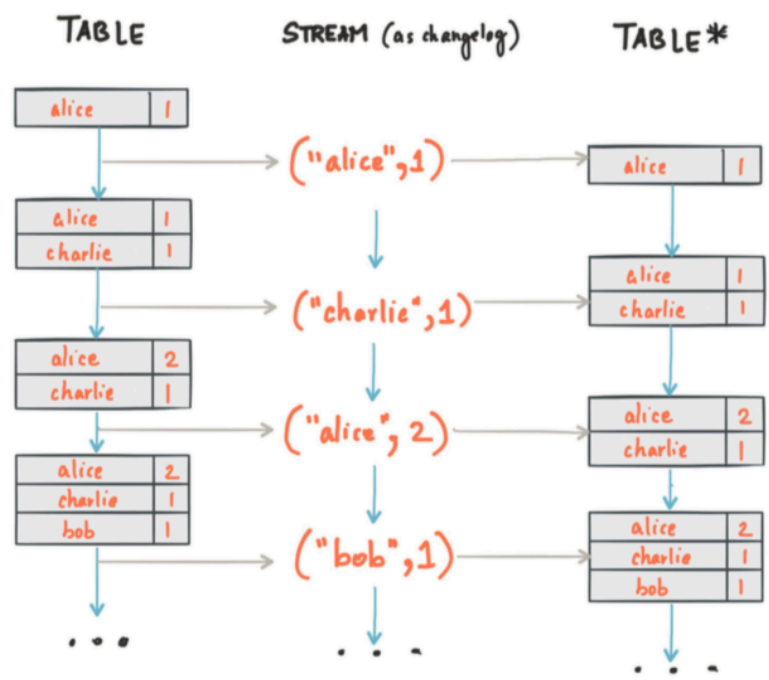


Figura 3: Dualità tra stream di eventi e database

### 4.3. Vantaggi

Event sourcing è un pattern estremamente utile per tutti quegli use-case dove è assolutamente necessario mantenere una storia dello sviluppo dello stato dell'applicazione del tempo. Tipici esempi sono gli strumenti di *versioning* del codice oppure la gestione di un storico bancario.

La capacità dell'event store di essere sia una struttura dati performante per la scrittura dei dati (è una semplice operazione di scrittura di complessità costante in fondo ad una sequenza) che una cronologia di tutti gli avvenimenti del sistema, permette una gestione degli errori estremamente semplice.

In un qualsiasi database relazionale se durante il normale utilizzo dell'applicazione avviene un errore logico che porta il database ad uno stato non corretto è necessario un *rollback* dell'intero database ad una data antecedente l'errore per correggerlo; tale processo è dispendioso e incline agli errori.

L'analisi del motivo dell'errore può non essere di facile realizzazione con un database relazionale a meno che non siano in uso i meccanismi di CDC: senza una cronologia delle transazioni può essere molto difficile risalire al motivo dell'errore e ad una possibile soluzione.

Diversamente nel caso dell'utilizzo di Event Sourcing, la gestione e l'analisi di un errore è estremamente semplice.

Nel caso in cui l'errore, che sarà sempre un evento, non ha generato un effetto "domino" sul sistema (ovvero l'errore non ha portato all'esecuzione di una catena di errori), una volta individuato è possibile pubblicare un evento "contrario" alla causa dell'errore in modo tale da cancellare l'apporto dell'errore sul sistema.

Nel caso contrario, ovvero il caso in cui l'errore ha generato una sequenza di errori, per ottenere lo stato corretto del sistema basterà ripetere l'esecuzione di tutti gli eventi del sistema escludendo quello che ha generato l'errore sulla base di dati.

È bene notare che con ES la gestione degli errori nello stato del sistema è strettamente legata all'atomicità e definizione degli eventi del sistema: una corretta (semplice) definizione degli eventi del sistema porterà ad una cronologia del sistema più chiara e comprensibile.

#### 4.4. Svantaggi

Event sourcing potrebbe non essere utile per una applicazione che richiede frequenti e continue query di richiesta sullo stato del sistema.

Come descritto in precedenza per ottenere lo stato corrente del sistema è necessario eseguire tutti gli eventi pubblicati sull'event store partendo da uno stato iniziale; se la nostra applicazione richiede di eseguire molte query di ricerca sullo stato corrente del database sarà quindi necessario calcolare lo stato del sistema *ogni volta che viene eseguita una nuova richiesta* (un esempio di richiesta sullo stato è la ricerca di tutti i record che presentano una particolare caratteristica).

Le modalità per risolvere questo problema sono determinate dal dominio e uso dell'applicazione che utilizza ES ma generalmente per ovviare a questa debolezza vengono realizzati degli *snapshot* dello stato dell'applicazione da utilizzare per l'esecuzione delle query di ricerca.

La frequenza di generazione ed aggiornamento di questi snapshot è *strettamente* legata al dominio applicativo dell'applicazione.

## 5. Apache Kafka: Ecosistema

### 5.1. Introduzione

*Publish/Subscribe* è un pattern architetturale utilizzato per la comunicazione asincrona tra diversi processi od oggetti.

In questo schema mittenti e destinatari dialogano tra loro per mezzo di un *broker*, un processo incaricato, da una parte, di ricevere messaggi da dei mittenti e dall'altra di consegnare gli stessi messaggi a dei destinatari.

I destinatari non conoscono i mittenti, ed i mittenti non si interessano di chi sono i destinatari: l'unico compito del mittente è quello di pubblicare dei messaggi sul broker, starà poi al destinatario il compito di abbonarsi (dall'inglese *subscribe*) al broker in modo da ricevere tutti i nuovi messaggi.

Questo pattern viene spesso utilizzato quando ci si trova ad avere più processi o servizi che generano delle metriche o dei dati i quali sono di vitale importanza per altrettanti servizi.

Una soluzione alternativa sarebbe creare dei canali dedicati tra produttori e consumatori ma questo non permetterebbe alla struttura di supportare un numero sempre più elevato di servizi od oggetti ed in un mondo dove è sempre più frequente l'utilizzo di microservizi e il logging di eventi e dati (Big Data) porterebbe ad un debito tecnologico elevato e difficile da correggere.

È in questo contesto che nasce Apache Kafka, una *streaming platform* basata su un append-only log utilizzato da dei *producer* per pubblicare dei messaggi consumati da dei *consumer*.

I messaggi pubblicati persistono nel tempo, sono leggibili deterministicamente da qualsiasi consumer e distribuiti all'interno del sistema secondo particolari logiche in modo da garantire protezione da crash e scalabilità del sistema.

## 5.2. Messaggi

L'unità dati fondamentali in Kafka è chiamata *messaggio* o *record*.

Ogni messaggio è suddiviso in *key* (opzionale) e *value* e possono essere di qualsiasi formato; Kafka non impone particolari standard riguardo i formati dei dati utilizzabili all'interno del sistema ma con lo scorrere del tempo *Avro* è diventato lo standard de facto (Kreps, 2015).

Il campo **key**, quando definito, è un byte array utilizzato come metadata per garantire un particolare ordinamento all'interno di un *topic*, un altro elemento fondamentale dell'architettura.

Nonostante l'unità dati fondamentale sia il messaggio piccoli gruppi di messaggi diretti verso lo stesso topic vengono generalmente raggruppati in *batches*.

La gestione dei messaggi in batches nasce per motivi di efficienza nel bilanciare throughput e latenza: a fronte di una latenza più alta per la consegna di un batch, vengono sprecate meno le risorse del sistema che altrimenti si ritroverebbe costretto a gestire l'overhead di consegna di un batch per ogni singolo messaggio.

## 5.3. Topic e partizioni

Un *topic* è un elemento utilizzato in Kafka per categorizzare una collezione di messaggi, e consiste in un unico stream di dati.

Un topic è suddiviso in *partizioni*, append-only logs sui quali vengono persisti i messaggi generati dai *producers* (Fig. 4).

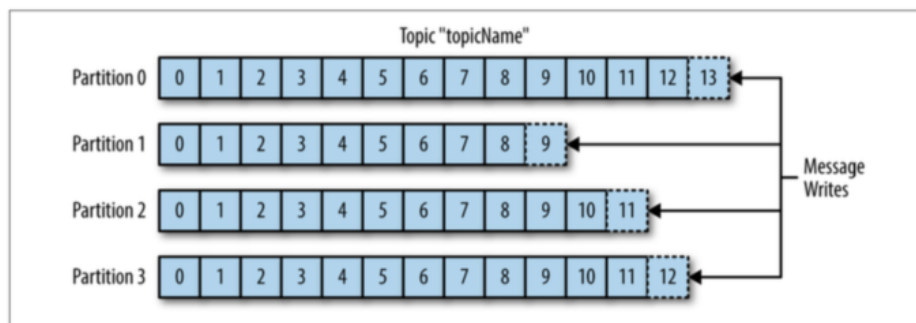


Figura 4: Un topic suddiviso in più partizioni

I messaggi sono inseriti in una partizione da un producer nell'ordine in cui sono stati inviati posizionandoli in fondo al log, non sono modificabili o cancellabili e sono contraddistinti da un *offset*, un indice numerico che funziona da *timestamp*

del messaggio. Un consumer legge i messaggi di un topic partendo dalla testa (o da uno specifico offset) del log proseguendo fino alla coda.

L'ordine di scrittura è garantito solo per ogni singola partizione: non è detto che messaggi appartenenti al medesimo topic siano in ordine cronologico se inseriti su partizioni diverse.

Per dare un esempio pratico di topic, supponiamo di utilizzare Kafka per creare uno *storage* di eventi ricevuti dal front-end di una applicazione: tipici eventi che vengono spesso loggati da un front-end possono essere *i link cliccati in una pagina, quali pagine sono state visualizzate in una sessione oppure se è stato visualizzato un video presente sulla pagina.*

Per ognuno di questi eventi verrà creato un singolo **topic** per raggruppare tutte le notifiche e dati generati da uno di quei particolari eventi a front-end: ad esempio avremo il topic **views-video-embeded** sul quale verranno registrati dei semplici **yes** o **no** con magari l'aggiunta di un **timestamp** (l'ora di visualizzazione), in questo modo il topic ci permetterà di calcolare la frequenza di visualizzazione del video.

## 5.4. Producers

Kafka è utilizzata da due tipologie di client: *producers* e *consumers*.

I *producers* hanno il compito di creare messaggi indirizzati a specifici topic (indipendentemente dal numero di partizioni da cui sono formati).

Come illustrato in precedenza, un topic è formato da un numero variabile di partizioni utilizzate come meccanismo di replica e gestione dei messaggi; alla creazione di un messaggio è possibile indicare al producer su quale partizione andare a scrivere il record specificando l'identificativo di una partizione specifica.

Il processo di pubblicazione di un messaggio (*Fig. 5*) inizia con la produzione di un **ProducerRecord** il quale deve contenere il topic sul quale vuole pubblicare il messaggio ad una value, ovvero il contenuto del messaggio; opzionalmente è possibile specificare una chiave o una partizione specifica.

Nella maggior parte dei casi d'uso di Kafka, il producer non si pone mai il problema di decidere su quale partizione andare a scrivere un particolare messaggio ma piuttosto vengono utilizzati dei meccanismi di *load-balancing* per spartire correttamente i messaggi su tutte le partizioni disponibili presenti nel topic.

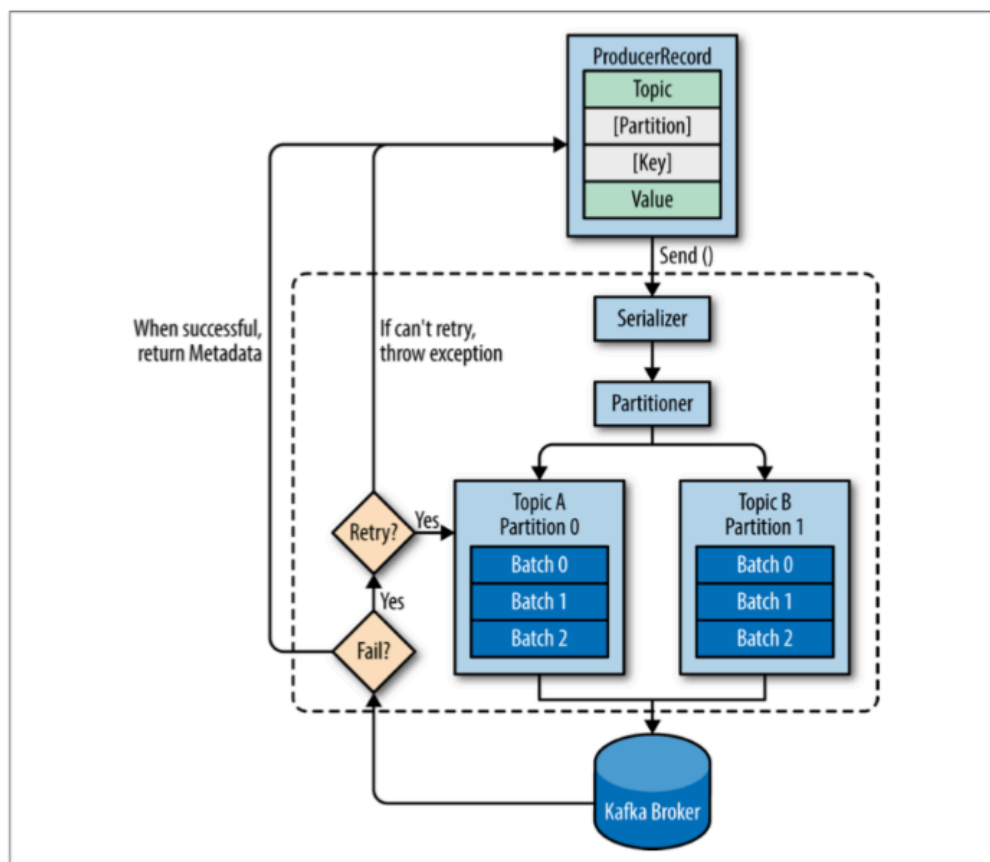


Figura 5: Processo di pubblicazione di un messaggio

Tipici esempi di algoritmi di load-balancing sono il calcolo della partizione in base ad una *hash key* derivata dall'offset del messaggio oppure utilizzando un algoritmo *round-robin*; se necessario è presente la possibilità di specificare un *partitioner* creato su misura al caso d'uso.

Una volta creato il **ProducerRecord** il producer serializza chiave e value del messaggio in **ByteArray** in modo da effettivamente trasmetterli sulla rete.

I dati sono quindi recapitati ad un *partitioner* che andrà a decidere su quale partizione pubblicare il messaggio (solo nel caso in cui la partizione non era già stata specificata nel record).

Il record viene quindi aggiunto ad un batch di record e il producer resta in attesa di avere abbastanza record (o alternativamente in attesa della scadenza di un *time-out*) prima di inviare il batch di messaggi ad un broker.



Una volta che il broker riceve il batch di messaggi verranno effettuati una serie di controlli per garantire la validità dei messaggi del batch rispetto al topic su cui si sta cercando di pubblicare questi messaggi.

In caso positivo il broker invia al producer un **RecordMetadata** con topic, partizione e offset dei messaggi dei pubblicati, altrimenti ritornerà un errore.

In caso negativo il producer può provare a rinviare il batch di messaggi.

Ogni partizione in un cluster ha un *leader* ed un insieme di sue *repliche* distribuite sui vari brokers.

Tutte le pubblicazioni dirette ad una particolare partizione devono prima essere pubblicate sul leader ed in un secondo momento devono essere replicate sulle repliche (o *followers*); questo meccanismo è utilizzato per garantire la durabilità dei dati in un cluster: se uno dei leader muore, viene eletta una delle repliche a nuovo leader della partizione e viene creata una nuova replica.

È possibile garantire diversi livelli di durabilità a seconda di come viene configurato il producer, favorendo od evitando problemi di scrittura e lettura dei messaggi a scapito del throughput.

Per poter creare un producer sono necessari tre parametri:

- **bootstrap.servers**: lista degli indirizzi (**host:port**) dei brokers del cluster Kafka che vogliamo utilizzare.
- **key.serializer**: nome della classe che verrà utilizzata per serializzare in byte array le chiavi dei record che vogliamo pubblicare con Kafka. È possibile crearne di nuovi implementando **org.apache.kafka.common.serialization.Serializer**.
- **value.serializer**: nome della classe che verrà utilizzata per serializzare in byte array il record da pubblicare.

Un esempio di producer è dato dal seguente codice Scala:

```

import java.util.Properties
import org.apache.kafka.clients.producer.{Callback, RecordMetadata}
import org.apache.kafka.clients.producer.{ProducerRecord, KafkaProducer}
import scala.concurrent.Promise

case class Producer(topic: String){
  val props = new Properties()
  props.put("bootstrap.servers", "localhost:9092")
  props.put("key.serializer",
    "org.apache.kafka.common.serialization.StringSerializer")
  props.put("value.serializer",
    "org.apache.kafka.common.serialization.StringSerializer")

  private val producer = new KafkaProducer[String, String](props)

  def send(value: String) = {
    val record = new ProducerRecord[String, String](topic, value)

    try {
      producer.send(record).get()
    } catch {
      case e: Exception => e.printStackTrace
    }
  }

  def sendAsync(value: String) = {
    val record = new ProducerRecord[String, String](topic, value)
    val promise = Promise[(RecordMetadata, Exception)]()

    producer.send(record, new Callback {
      override def onComplete(metadata: RecordMetadata,
        exception: Exception) = {
        promise.success((metadata, exception))
      }
    })
  }
}

```

Creato un oggetto **Properties** con le configurazioni di base, viene creato un nuovo producer capace di pubblicare dei record (di tipo **String**) ad un topic (passato come parametro alla case class) con il metodo **send(value: String)**.

Un producer può inviare record con tre diverse modalità:

- **Fire-and-forget**: il messaggio viene mandato al server e viene ignorata la possibilità che questo messaggio non venga ricevuto.
- **Sincrona**: il messaggio viene mandato con un metodo **send()** il quale restituisce un **Future**, viene quindi utilizzato il metodo **.get()** per attendere una risposta dal server per capire se il messaggio è effettivamente stato ricevuto.
- **Asincrona**: il metodo **send()** restituisce una **callback** che verrà eseguita quando (e se) riceverà una risposta dal broker Kafka.

E' possibile configurare un producer secondo una moltitudine di campi di configurazione, ma sicuramente uno dei più importanti è il valore attribuito al campo **acks** il quale è direttamente collegato alla durabilità dei messaggi e definisce quante repliche devono ricevere il messaggio pubblicato sul leader prima di poter garantire al producer la corretta pubblicazione del messaggio.

Esistono tre possibili valori per **acks**:

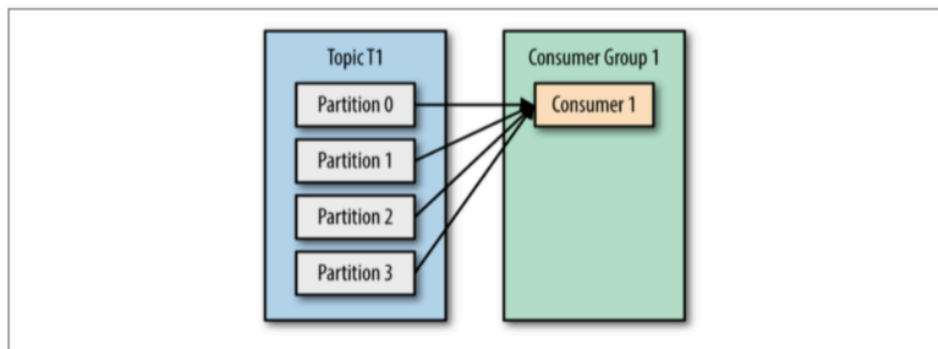
- con **acks=0**, il producer non si aspetterà di ricevere un messaggio di conferma da parte del broker. Questo implica che nel caso di un malfunzionamento, il producer non sarà a conoscenza del fallimento ed il messaggio verrà perso. È l'opzione che garantisce il più alto valore di throughput.
- con **acks=1**, il producer riceverà un messaggio di conferma di pubblicazione del messaggio appena *almeno* una replica confermerà di aver ricevuto il messaggio pubblicato sul leader. Con questa opzione è comunque possibile perdere il messaggio se, a seguito della morte del leader, viene eletto come nuovo leader non la replica che aveva confermato la ricezione del messaggio ma piuttosto una delle repliche che non lo avevano ancora ricevuto.
- con **acks=all**, il producer riceverà conferma della pubblicazione del messaggio solo dopo che tutte le repliche hanno ricevuto il messaggio pubblicato sul leader. Questa opzione garantisce la pubblicazione di un messaggio a scapito di un minor throughput.

## 5.5. Consumers

Un *consumer* è un client Kafka utilizzato per leggere dati da un topic e tipicamente è parte di un *consumer group*.

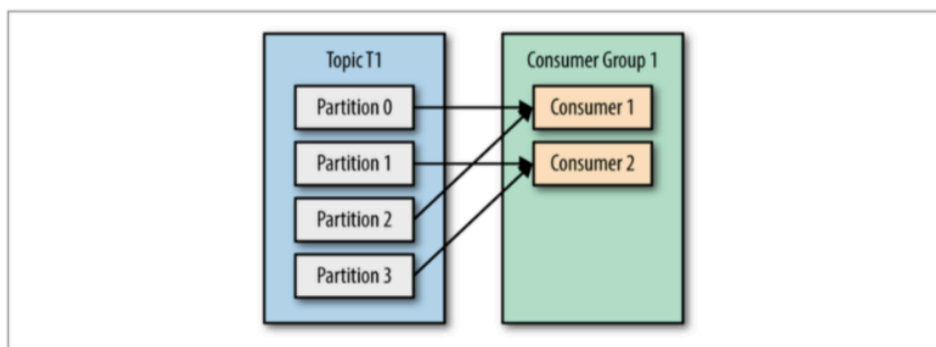
Un consumer group è definito da uno specifico **group.id** ed un topic che tutti i membri del gruppo hanno in comune; ogni partizione del topic è letta da un solo consumer del gruppo e più consumer del gruppo non possono leggere dalla stessa partizione.

Supponiamo di avere un topic T1 con quattro partizioni (*Fig. 6*), nel caso in cui il nostro gruppo è formato da un solo consumer C1 sarà solo lui a consumare l'intero topic.



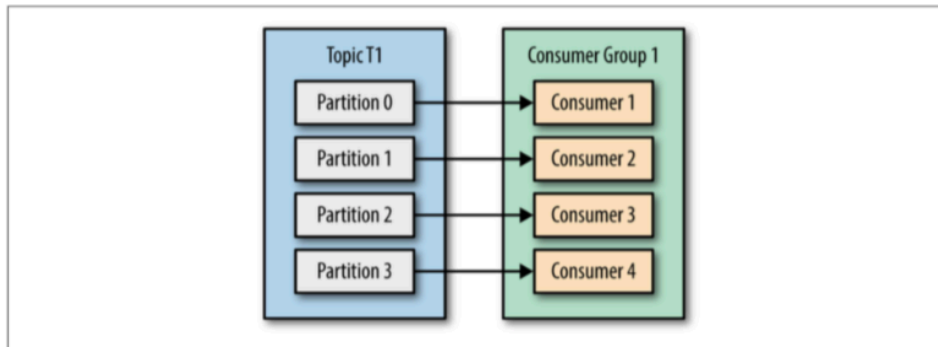
*Figura 6: Un consumer con quattro partizioni*

Se aggiungiamo un nuovo consumer C2 al gruppo (*Fig. 7*), ogni consumer riceverà dati da solo due partizioni disponibili.



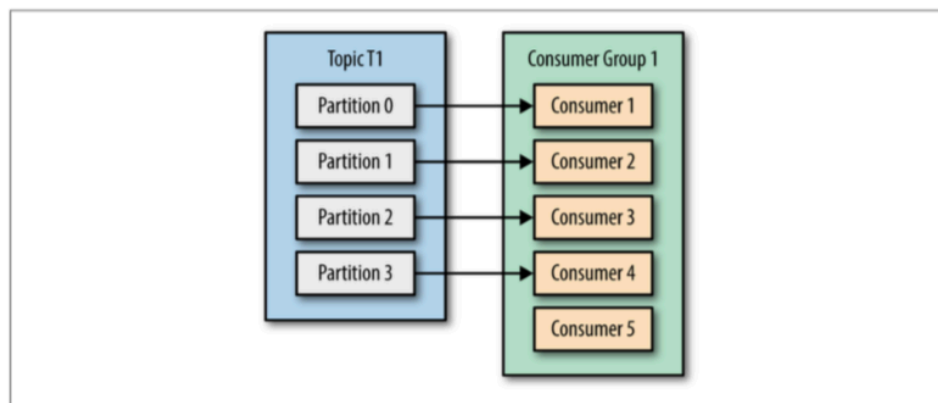
*Figura 7: Due consumer consumano quattro partizioni*

Con quattro consumer, ogni consumer leggerà da una sola partizione (*Fig. 8*).



*Figura 8: Quattro consumer consumano l'intero topic individualmente*

Ed infine, nel caso in cui il gruppo contenga più consumer del numero di partizioni del topic vi saranno sicuramente dei consumer che non leggeranno dal topic (*Fig. 9*).



*Figura 9: Numero di consumers superiore al numero delle partizioni*

Questo meccanismo di bilanciamento dei consumer rispetto alle partizioni di un topic permette di scalare l'architettura orizzontalmente nel caso di topic con grossi moli di dati sensibili che necessitano di essere consumati da applicazioni con operazioni ad alta latenza come la scrittura su un database o l'esecuzione di calcoli: con Kafka, per evitare situazioni a "collo di bottiglia", basta aumentare il numero di partizioni di un topic ed il numero di consumer di quel particolare topic.

Un altro importante vantaggio dell'utilizzo dei consumer group è dato dalla possibilità di ribilanciare il gruppo nel caso di crash, morte o aggiunta di un consumer.

Per *ribilanciare il consumer group* si intende il processo di cambio di proprietà di una partizione da un consumer A ad un consumer B.

Questo procedimento insieme alla possibilità di aggiungere e rimuovere consumer per ogni partizione, è ciò che permette a Kafka di scalare la propria architettura su grossi numeri di record e topic. È importante notare che questa funzionalità non è comunque desiderabile: durante un ribilanciamento il consumer group non può consumare i dati di un topic, comportando quindi un rallentamento nella lettura dei dati.

Un consumer per non risultare morto deve inviare degli *heartbeats* al broker eletto a *coordinator* del consumer group. Questi "segni di vita" sono inviati al broker ogni volta che il consumer tenta di leggere dal topic.

Nel caso in cui il broker non riceva un heartbeat da un consumer entro un particolare lasso di tempo, verrà subito scatenato un ribilanciamento del gruppo a cui appartiene il consumer. Allo stesso modo, nel caso in cui un consumer venga rimosso dal gruppo, sarà lo stesso consumer ad inviare un messaggio di "uscita" dal gruppo al broker che anche in questo caso forzerà un ribilanciamento del consumer group.

Un esempio di consumer sviluppato in Scala è dato dal seguente codice:

```
import java.util.Properties
import org.apache.kafka.clients.consumer.KafkaConsumer

case class Consumer(topic: String){
  val props = new Properties()
  props.put("bootstrap.servers", "localhost:9092")
  props.put("key.deserializer",
    "org.apache.kafka.common.serialization.StringSerializer")
  props.put("value.deserializer",
    "org.apache.kafka.common.serialization.StringSerializer")
  props.put("group.id", "example")

  private val consumer = new KafkaConsumer[String, String](props)
  consumer.subscribe(util.Collections.singletonList(topic))

  def readTopic(){
    while(true){
      val records = consumer.poll(100)
      for (r <- records.asScala){
        println(s"${r.offset} ${r.key} ${r.value}")
      }
    }
  }
}
```

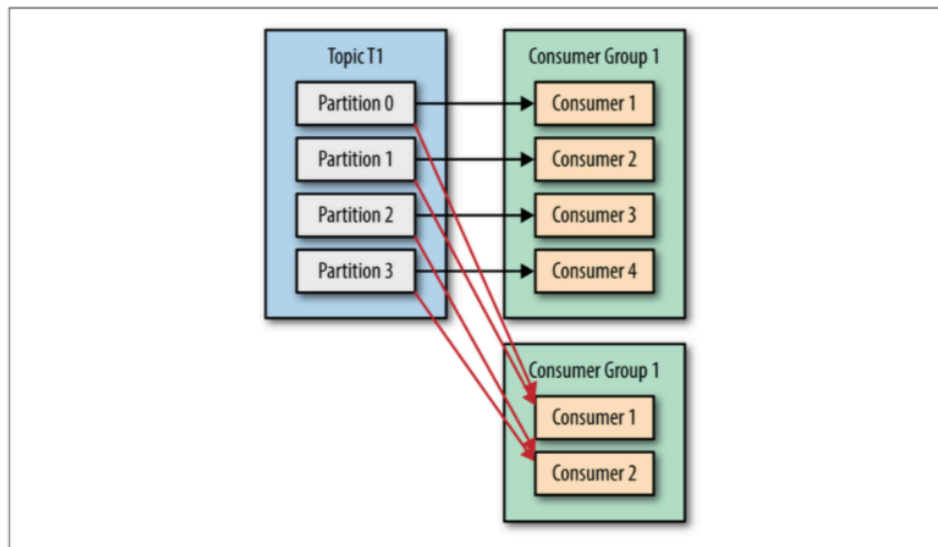
Come in precedenza con l'implementazione del producer, prima di poter creare un producer è necessario definire alcune configurazioni minime:

- **bootstrap.servers:** l'indirizzo del broker
- **key.deserializer:** il tipo di classe da utilizzare per deserializzare la chiave dei dati pubblicati su un topic
- **value.deserializer:** il tipo di classe da utilizzare per deserializzare i dati pubblicati su un topic
- **group.id:** l'identificativo del consumer group di cui il consumer fa parte

I consumer leggono i dati di un topic attraverso un meccanismo di pull ovvero sono loro stessi a richiedere al broker i dati.

Ogni volta che un consumer decide di leggere da un topic è lui stesso a tenere traccia dell'ultimo messaggio che è stato letto e per tenerne traccia utilizzerà un *offset*: un identificativo corrispondente alla posizione del messaggio nella partizione (il primo messaggio avrà offset pari a 0, l'n-esimo messaggio offset n, ecc.).

Dato che sono gli stessi consumer a tenere traccia dell'offset dei messaggi letti e che sono sempre loro a richiedere i dati al broker consumer group diversi possono leggere dallo stesso topic (*Fig. 10*) senza che questo perda alcun messaggio (i record presenti in un topic sono immutabili).



*Figura 10: Un topic con più gruppi di consumer*

La lettura di una partizione può partire o dal primo messaggio esistente oppure specificando un particolare offset.

Come si può vedere dalla funziona **readTopic()** definita nell'esempio, un consumer utilizza un loop infinito di chiamate a **.poll()** definito su un intervallo di tempo variabile (ovvero il metodo messo a disposizione dall'interfaccia di Kafka) per richiedere al broker tutti i messaggi che quel consumer non ha ancora letto.

Per sapere quali messaggi inviare il broker deve ricevere dal consumer l'offset dell'ultimo messaggio letto attraverso una operazione di *commit*; gli offset vengono pubblicati dai consumer su di un particolare topic chiamato **\_\_consumer\_offset** al quale tutti i broker hanno accesso e a cui fanno riferimento per calcolare quali messaggi inviare.

Il topic **\_\_consumer\_offset** è inoltre utilizzato nel caso di ribilanciamento di un gruppo. A seguito di un ribilanciamento un o più consumer del gruppo possono ricevere un nuovo insieme di partizioni, e per capire da dove iniziare il nuovo processo di lettura, devono interrogare il topic **\_\_consumer\_offset**.

Se l'offset pubblicato su **\_\_consumer\_offset** è *minore* dell'offset dell'ultimo messaggio che il consumer client ha in memoria, tutti i messaggi compresi tra i due offset verranno riprocessati (*Fig. 11*).

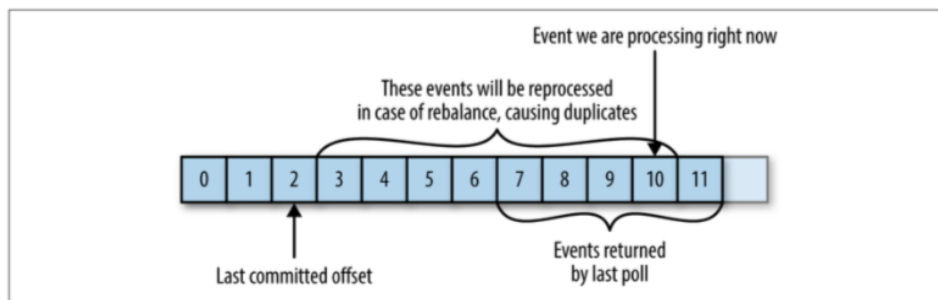


Figura 11: Replay dei messaggi in base all'offset

Se l'offset pubblicato su **\_\_consumer\_offset** è *maggiore* dell'offset dell'ultimo messaggio che il consumer client ha in memoria, tutti i messaggi compresi tra i due offset verranno **persi** (*Fig. 12*).

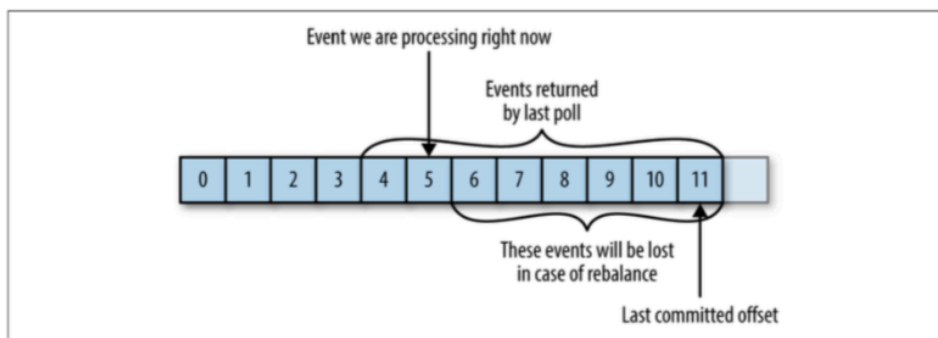


Figura 12: Perdita di messaggi in base all'offset



Per una corretta gestione dei messaggi è quindi di assoluta importanza la capacità di gestire gli offset in modo adeguato alle necessità di ogni progetto, per questo motivo esistono varie possibile configurazione del processo di *commit*:

- **automatic commit**: ogni cinque secondi il consumer genera un commit dell'ultimo offset letto con `.poll()`, è la configurazione di default di ogni consumer.
- **commit current offset**: è lo stesso consumer a decidere quando inviare l'ultimo offset letto attraverso l'uso della funzione `commitSync()`: alla chiamata della funzione viene inviato al broker l'ultimo offset letto da `.poll()` ed il consumer rimane in attesa di un segnale di **acknowledgment** da parte del broker, in caso positivo il consumer continuerà nel processo di lettura, altrimenti verrà lanciata un'eccezione. È prevista la possibilità di riprovare un numero di volte il processo di commit.
- **commit current offset in modalità asincrona**: come la modalità precedente ma non bloccante per il consumer, l'offset viene inviato chiamando `.commitAsync()`.

## 5.6. Brokers e clusters

Un *broker* è un server Kafka con svariati compiti quali ricevere, indicizzare e salvare i messaggi inviati dai producers ed inviare i messaggi richiesti dai consumers; un singolo broker è capace di gestire migliaia di partizioni e milioni di messaggi al secondo.

I broker sono stati creati per lavorare in *clusters* (Fig. 13) ovvero gruppi di brokers ordinanti secondo una particolare gerarchia.

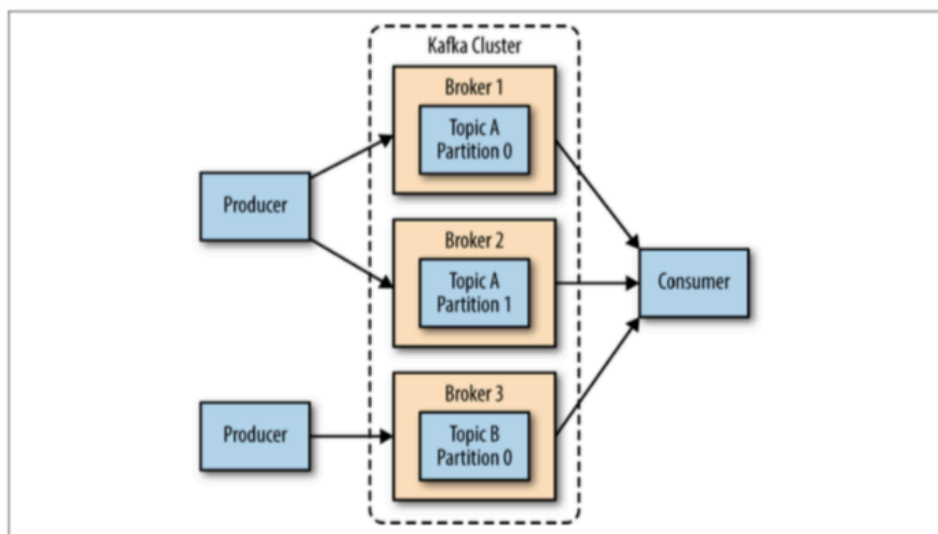
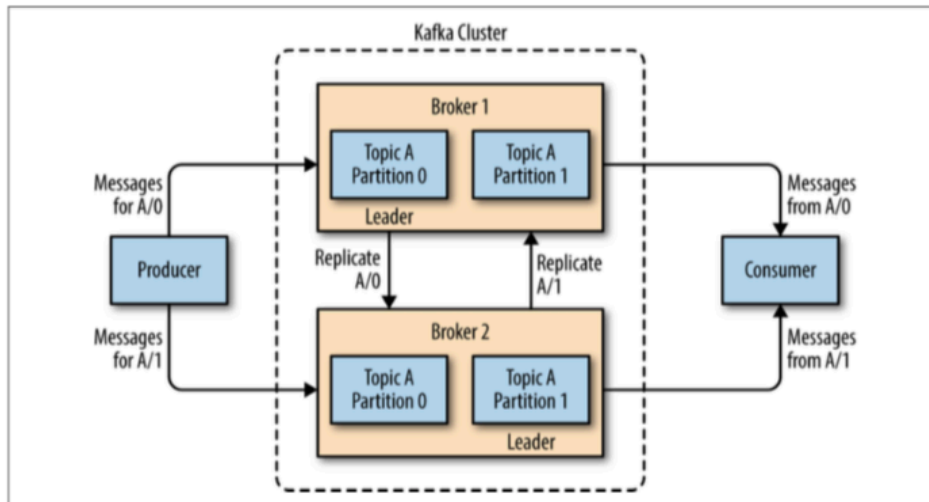


Figura 13: Cluster

A capo di un cluster troviamo un broker *leader* al quale tutti gli altri broker del cluster devono far riferimento per permettere ai meccanismi di replicazioni dei messaggi di funzionare correttamente: una partizione può essere assegnata a più broker, questo permette al cluster di poter gestire fallimenti dei brokers (*Fig. 14*).

In ogni cluster un particolare broker viene eletto a *controller*, ovvero un broker con l'incarico di gestire la suddivisione delle partizioni sull'intero cluster e di monitorare l'andamento del cluster.



*Figura 14: Gestione delle repliche*

Una funzionalità importante di Kafka è la possibilità di utilizzare i topic come database di messaggi persistenti.

I messaggi vengono tenuti in memoria per un particolare periodo di tempo oppure in base allo spazio di memoria di occupato, entrambe le opzioni sono configurabili alla creazione di un broker, vi è poi la possibilità di abilitare la **log compaction** ovvero un meccanismo che permette a Kafka di mantenere in memoria *solo gli ultimi messaggi indicizzati su un indicativo specifico*.

## 5.7. Schema

Nonostante i messaggi in Kafka non siano altro che degli array di byte è fortemente consigliato l'uso di *schema* per la gestione e l'uso della struttura dei record.

Lo *schema* è la struttura o organizzazione logica dei dati contenuti in un topic (Neha Narkhede, 2017) e nel caso specifico di Kafka, la scelta dei formati disponibili ricade spesso su di un singolo formato: Apache Avro.

Esistono altre scelte possibili come Javascript Object Notation (JSON) oppure Extensible Markup Language (XML), ma Avro offre una serie di vantaggi rispetto a questo genere di schemi oltre ad avere alcune implementazioni ad-hoc in Kafka.

Avro è diventato nel tempo lo standard per gli, gli stessi sviluppatori di Kafka ne promuovono l'uso citando una serie di motivi:

- Avro è mappabile su JSON
- Al contrario di JSON, è possibile scomporre lo schema dei dati dalla definizione dell'oggetto
- È un linguaggio maturo ben supportato dalla community: Esistono molte librerie che permettono di creare automaticamente oggetti Java o case classes in Scala partendo da uno schema Avro

Apache Avro è un formato per la serializzazione dei dati: ogni messaggio Avro si divide in due parti: i *dati* ed il loro *schema*.

In *Fig. 16* è illustrato un esempio di un generico record dati definito secondo lo schema AVRO in *Fig. 15*.

```
{
  "type": "record",
  "doc": "This event records the sale of a product",
  "name": "ProductSaleEvent",
  "fields" : [
    { "name": "time", "type": "long", "doc": "The time of the purchase" },
    { "name": "customer_id", "type": "long", "doc": "The customer" },
    { "name": "product_id", "type": "long", "doc": "The product" },
    { "name": "quantity", "type": "int" },
    { "name": "payment",
      "type": { "type": "enum",
        "name": "payment_types",
        "symbols": [ "cash", "mastercard", "visa" ] },
      "doc": "The method of payment" }
  ]
}
```

Figura 15: Schema in formato AVRO

```

{
  "time": 1424849130111,
  "customer_id": 1234,
  "product_id": 5678,
  "quantity": 3,
  "payment_type": "mastercard"
}

```

Figura 16: Record di dati

In Kafka ogni topic è definito da uno schema AVRO in modo da:

- definire la struttura dei messaggi pubblicabili nel topic
- permettere a producers e consumers di conoscere quali sono i campi dei messaggi del topic e qual è il loro tipo
- documentare la tipologia dei messaggi pubblicati nel topic
- evitare la presenza di dati corrotti nel topic

Questo genere di meccanismo per la gestione dei dati diventa di assoluta importanza all'aumentare delle applicazioni che dipendono dall'utilizzo dei dati prodotti e gestiti da una piattaforma Kafka, evitando problemi "effetto Domino" dove un singolo errore in un messaggio potrebbe portare a corrompere un consumer o applicazioni terze che utilizzano i dati forniti dal consumer.

Un formato dei dati consistente come AVRO permette di *disaccoppiare i formati utilizzati per la lettura e la scrittura dei messaggi*: viene data la possibilità alle applicazioni che si iscrivono ad un particolare topic di poter utilizzare un nuovo schema di dati compatibile con un vecchio formato, senza dover aggiornare tutte le applicazioni che utilizzano ancora il vecchio formato.

Supponiamo ad esempio di utilizzare un formato del tipo seguente per gestire le informazioni riguardando agli acquirenti di un particolare servizio/piattaforma:

```

{
  "namespace": "customerManagement.avro",
  "type": "record",
  "name": "Customer",
  "fields": [
    {"name": "id", "type": "int"},
    {"name": "name", "type": "string"},
    {"name": "faxNumber", "type": ["null", "string"], "default": "null"}
  ]
}

```

Una applicazione che vuole utilizzare lo stream di dati di questo topic avrà probabilmente dei metodi come `getId()`, `getName()` e `getFaxNumber()` per leggere i dati del topic; di nota è il tipo del campo `faxNumber` il quale è esplicitamente

possibile che sia **null**, ovvero è lecito aspettarsi che l'applicazione che utilizza questi dati non si romperà nel caso in cui il fax non sia presente nel messaggio.

Supponiamo di aver utilizzato lo schema precedente per generare una importante mole di dati in un topic ma di voler migrare il nostro schema ad un nuovo formato che permetta agli utenti di specificare la loro **email** piuttosto che il loro fax.

```
{
  "namespace": "customerManagement.avro",
  "type": "record",
  "name": "Customer",
  "fields": [
    {"name": "id", "type": "int"},
    {"name": "name", "type": "string"},
    {"name": "email", "type": ["null", "string"], "default": "null"}
  ]
}
```

Ancora una volta l'applicazione che utilizza questo schema avrà dei metodi **getId()**, **getName()** ma invece di **getFaxNumber()** avrà **getEmail()**.

Dopo l'aggiornamento dello schema, i vecchi messaggi presenti nel topic conterranno il campo **faxNumber** mentre i nuovi messaggi avranno il campo **email**.

Dato che il tipo dei campi **faxNumber** e **email** può essere sia **string** che **null**, nessuna delle tue tipologie di applicazioni potrà fallire: la vecchia tipologia di applicazioni semplicemente registrerà i nuovi messaggi come utenti senza un numero di fax, mentre le nuove applicazioni vedranno i vecchi messaggi del topic come utenti senza una email.

Questo genere di *evoluzione* dello schema dei dati di un topic è il motivo centrale dietro all'uso della tecnologia: garantire la robustezza dei dati senza compromettere la leggibilità dello schema o il funzionamento di applicazioni che utilizzano lo stesso topic.

L'evoluzione dello schema è permessa solo secondo determinate regole di compatibilità la cui definizione esula dal contesto di questo tesi ma che possono essere visionate nella [documentazione di Apache Avro](#).

## 5.8. Schema Registry

Uno dei vantaggi di Avro rispetto a JSON è la possibilità di non dover inserire lo schema dei dati completo in ogni record permettendo di pubblicare su un topic dei messaggi meno pesanti ed è proprio per sfruttare questo vantaggio che nasce lo *Schema Registry*.

Uno *schema registry* è un servizio di gestione degli schema Avro utilizzato da Kafka per l'evoluzione e la gestione dei metadati di un topic (Neha Narkhede, 2017); è utilizzato dai producers nella fase di scrittura (serializzazione di un messaggio), dai consumer nella fase di lettura (deserializzazione di un messaggio) (Fig. 17) ed impone delle regole di forma a tutti i client che intendono utilizzare un topic specificato con formato dati Avro.

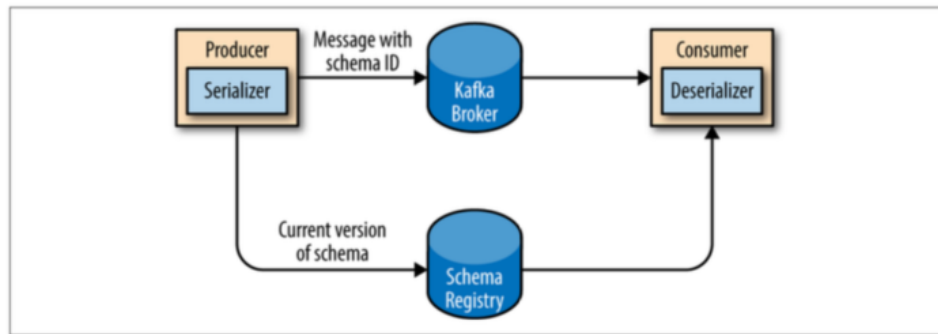


Figura 16: Serializzazione e deserializzazione con schema registry

Supponiamo di avere uno schema registry e di voler produrre dei messaggi su di un topic:

1. Il producer interrogherà il registry per sapere se esiste già uno schema dei dati per un particolare topic inviando la propria copia dello schema ed in caso contrario sarà lui stesso a pubblicarlo nel registry.
2. Lo schema registry verifica se lo schema ricevuto dal producer è uguale o una evoluzione compatibile dello schema già presente, in caso negativo verrà alzata un'eccezione e vietata la scrittura al producer.
3. Se lo schema proposto dal producer è valido nel record Avro verrà inserito un riferimento allo schema del topic.

L'utilizzo dello schema registry da parte di un consumer è speculare a quello di un producer.

## 6. Architetture Event-driven

### 6.1. Kafka come piattaforma per Event Sourcing

Il pattern Event Sourcing e la piattaforma Apache Kafka sono accomunati da uno specifico elemento: *il log*.

Entrambe le soluzioni utilizzano la medesima struttura dati per risolvere un particolare problema, è quindi spontaneo che nel tempo ES sia diventato un pattern predominante nelle applicazioni dell'ecosistema Kafka (Neha, Event sourcing, CQRS, stream processing and Apache Kafka: What's the connection?, 2016).

Riassumendo i capitoli precedenti Event Sourcing utilizza il log (*Fig. 18*) come struttura dati per poter mantenere una sequenza immutabile di eventi generati da una applicazione o da più parti di un sistema; questa sequenza di eventi definisce la storia del sistema e può essere utilizzato per ricavare lo stato del sistema in un qualsiasi momento della sua vita.

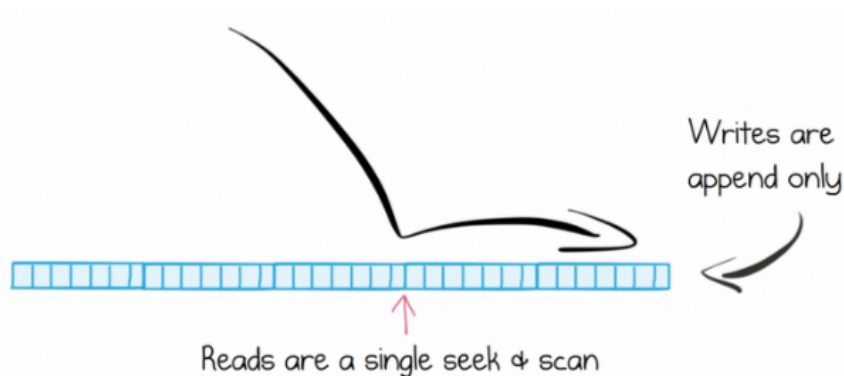


Figura 17: Scrittura e lettura da un log

Il log è una struttura dati fondamentale e naturalmente efficiente dato il caso d'uso: sia la lettura che la scrittura di dati da un log (o topic Kafka) sono sequenziali e sfruttano a pieno le funzionalità di *batching* e *caching* offerte da Kafka.

Nel caso di un producer l'operazione di inserimento di un nuovo elemento nel topic è estremamente efficiente in quanto si tratta di un semplice inserimento in coda di costo costante e nel caso di utilizzo di un consumer la lettura da un topic sarà sempre effettuata in batch e sequenziale (che sia dall'inizio del topic o da un particolare offset è irrilevante) e quindi anche in questo caso avrà costo costante.

Il limite imposto dalla lettura sequenziale di un topic permette inoltre una riduzione dei costi di gestione della piattaforma non dovendo gestire in memoria indici tipici di altre strutture dati come tabelle hash o alberi utilizzati per letture

indicizzate o cancellazioni. La lettura sequenziale di un topic è pur sempre un limite della piattaforma ma che nel caso d'uso di una applicazione sviluppata seguendo event sourcing diventa assolutamente irrilevante in quanto è lo stesso pattern a premere per l'utilizzo di un log come event store.

L'utilizzo di Kafka come Event Store è un altro dei motivi che rende la piattaforma un ottimo strumento da utilizzare con Event Sourcing. Perché una piattaforma possa funzionare come database/event store è necessario che sia *affidabile*, *modulare* e dotata di strumenti per la creazione di *viste* degli eventi registrati.

Kafka è *modulare* grazie al sistema dei brokers il quale permette uno sviluppo dei consumer orizzontale, performante, "elastico" e facilmente ridimensionabile a seconda della mole di dati presente nel sistema.

Kafka è inoltre *affidabile* su più fronti:

- nonostante non sia una qualità necessariamente utile in un sistema che utilizza ES la *temporalità* dei messaggi è garantita attraverso la definizione di una chiave al momento della pubblicazione dei messaggi oppure tramite l'utilizzo di una singola partizione.
- la *durabilità* dei messaggi è garantita dal meccanismo di replica su più partizioni.
- l'utilizzo dello schema registry garantisce maggiore sicurezza nel caso di *evoluzione dello schema dei dati*, rendendo disponibile uno strumento che automaticamente valida nuove evoluzioni della forma dei dati aiutando gli sviluppatori nel creare soluzioni backward-compatible (Shapira, 2015).
- il log con la sua qualità di *immutabilità* oltre ad essere la struttura dati dominante della piattaforma ne è anche il *sistema di versioning dei dati* garantendo l'esistenza di una *cronologia non modificabile* di tutti gli eventi che sono stati registrati.
- la possibilità di creare un numero arbitrario di topic permette la *segmentazione* dei dati/eventi secondo la granularità che il team di sviluppo ritiene più opportuna; non è assolutamente sbagliato sviluppare una soluzione con microservizi che utilizzano topic "privati" per le loro necessità e topic "pubblici" per comunicare tra loro (*Fig. 19*).
- Kafka permette di specificare per quanto tempo mantenere l'intero stato dei propri topic rendendo di fatto la piattaforma un "database per eventi".
- attraverso l'uso di *Kafka Connect API* è estremamente facile leggere e scrivere dati da/a database *esterni*.
- per la lettura e pubblicazione di dati da/a Kafka a microservizi *Kafka Streams API* offre una soluzione semplice e ad-hoc basata su event store.



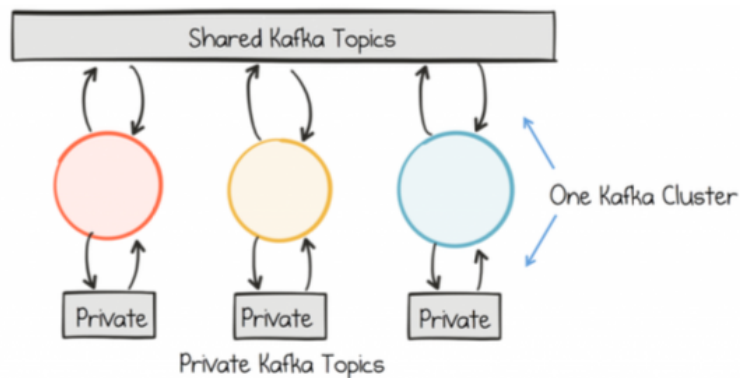


Figura 19: Topic pubblici e privati in un sistema distribuito

Il pattern Event Sourcing si basa sull'idea di avere una sequenza di eventi che, ove necessario, è possibile eseguire per ottenere lo stato corrente del sistema.

In genere la costruzione dello stato attuale della piattaforma è una operazione dispendiosa e ES non dovrebbe essere preso in considerazione se l'unico approccio all'analisi o uso dei dati del sistema è una serie di query allo stato attuale, ma è anche vero che avere una vista aggiornata dello stato corrente può essere utile anche per quei sistemi che decidono di abbracciare completamente l'approccio basato sugli eventi.

I topic in Kafka sono dei log di eventi immutabili ma con una particolarità: è possibile *comprimerli* in base al valore della chiave (Fig. 20).

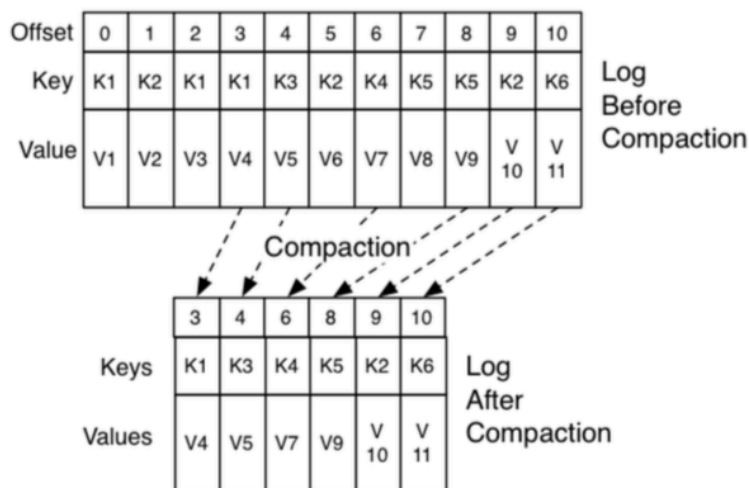


Figura 20: Log compaction

Il sistema di *log compaction* garantisce che Kafka, dato un topic, conserverà solamente l'ultimo messaggio a parità di chiave pubblicato sullo stesso topic. Questa funzionalità può quindi essere utilizzata come *vista dello stato corrente del*

*sistema* o di alcune sue parti oppure come *meccanismo di difesa* per la fase di recovery post-crash del sistema.

Il problema del calcolo del stato corrente è quindi risolto affiancando ai log utilizzati come stream di eventi un topic compresso "istanza" del stato del sistema.

È da notare che questo meccanismo di compressione dei log può anche essere utilizzato per archiviare parte di un vecchio log in modo da risparmiare spazio in memoria.

## 6.2. Kafka Connect: sistemi legacy, database ed Event Sourcing

Se si decide di sviluppare un nuovo progetto partendo da zero ed utilizzando Kafka ed ES non è difficile immaginare che creare la nuova struttura sarà relativamente semplice in quanto ci ritroveremo a sviluppare soluzioni software basate su eventi e verrà naturale pubblicare questi eventi su Kafka. Nel caso in cui si decida di adottare Kafka ed ES su un vecchio progetto sorgono spontanee domande come "Quali sono i miei eventi?", "Come posso utilizzare i database già presenti?" oppure "Ultimata la migrazione che fine faranno i vecchi database e le vecchie applicazioni?".

Domande come queste sono risolte dalla *Connect API* di Kafka.

Kafka Connect è un framework che permette di integrare Kafka con moltissime delle soluzioni database già presenti sul mercato (MySQL, PostgreSQL, Elasticsearch, Mongo, ecc.). Il framework si basa sull'uso di *Source connectors* per importare dati da sorgenti esterne e *Sink connectors* per la fase di scrittura su strutture esterne; Kafka viene rilasciato di base con molti connettori già pronti e testati all'uso e molti altri sviluppati dalla community open-source sono disponibili online. È inoltre prevista la possibilità di sviluppare connettori custom per soluzioni più particolari del normale.

I source connector funzionano facendo leva sul log transazionale generato tramite Change Data Capture (CDC) dal database sorgente che non è altro che un log sul quale vengono scritte, in ordine temporale, tutte le operazioni di insert, update e delete fatte sul database in questione. È facile immaginare che questo log transazionale è utilizzabile per definire lo stream di eventi necessario a Kafka per funzionare ed è esattamente così che funzionano i connettori: definita una configurazione per il connettore il transactional log del database viene copiato su di un topic Kafka. Esistono ovviamente delle differenze nel funzionamento di alcuni connettori (alcuni database supportano CDC basato su polling, altri su meccanismi push-pull, ecc.) ma il risultato del procedimento di creazione del topic legato al database sorgente sarà sempre lo stesso.

Specularmente i sink connector sono invece utilizzati per operazioni di write/update su dei database partendo da un topic specificato.

Di nota è la possibilità, grazie ai sink connectors, di utilizzare svariati database come data store per topic compressi garantendo così al sistema la possibilità di sfruttare le classiche tecnologie basate su database per tutte quelle operazioni di query dei dati per il quale Kafka non è necessariamente la soluzione migliore.

Utilizzando source e sink connectors è quindi possibile creare pipeline complesse di dati tra Kafka, vari database e architetture "legacy" in modo da permettere al sistema di evolversi gradualmente verso una soluzione basata solamente su Kafka (se necessario) (Stopford, Using Apache Kafka as a Scalable, Event-Driven Backbone for Service Architectures, 2017).

Si immagini ad esempio di utilizzare i vecchi dati presenti in una applicazione legacy per popolare i topic della piattaforma Kafka e parallelamente sviluppare nuove soluzioni basate sulla nuova architettura per raggiungere gli utenti della applicazione (Fig. 21).

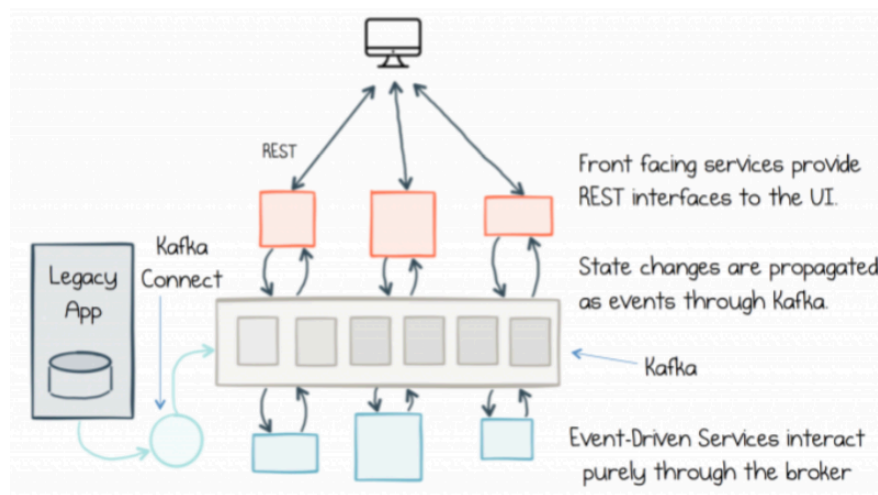


Figura 21: Architettura legacy con servizi Kafka

Un esempio potrebbe essere quello di un e-commerce: potremmo immaginare di avere il servizio di gestione e creazione dell'inventario come la nostra "legacy app" nell'immagine, mentre potremmo avere il servizio di validazione dell'ordine (conferma dell'ordine, informazioni di *tracking*, ecc.) gestito tramite Kafka.

### 6.3. Kafka Streams

Per poter sviluppare soluzioni software basate sullo streaming di eventi su topic è necessario trovare una soluzione per poter avere accesso ai dati in un topic in un formato modulare e facilmente utilizzabile da qualsiasi microservizio che decida di collegarsi allo stream di eventi (Stopford, Building a Microservices Ecosystem with Kafka Streams and KSQL, 2017). Questa soluzione è data dalla libreria/API *Kafka Streams*.

Kafka Streams è una libreria che fornisce una astrazione sul concetto di state store proposto da Kafka in modo da creare applicazioni o microservizi con uno qualsiasi dei linguaggi supportati (Java o Scala). Uno state store è un database chiave-valore creato per interfacciarsi con un topic Kafka; è possibile leggere e scrivere da un topic Kafka semplicemente utilizzando le funzioni proposte dalla libreria. La libreria offre delle primitive per creare delle applicazioni basate sul concetto di *topologie* ovvero un grafo di stream processors (i nodi del grafo) connessi tra loro da stream di dati (gli archi del grafo).

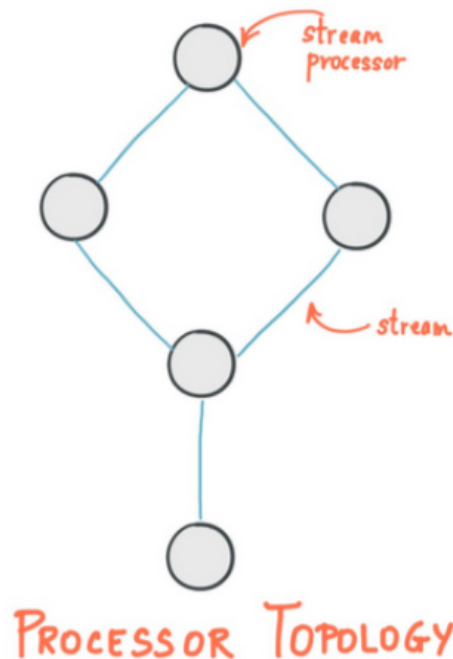


Figura 22: Generica topologia di una applicazione Kafka Streams

Uno *stream processor* è un nodo della topologia con il compito di processare in qualche modo i dati che riceve. Ogni processor riceve dal nodo precedente nella topologia un record alla volta, vi applica una operazione e produce uno o più nuovi record in output verso tutti i suoi successori.

Le operazioni eseguite da ogni nodo possono essere espresse utilizzando una Domain-specific Language (DSL) dichiarativa e funzionale (esempi tipici sono funzioni come **map** e **filter**) oppure una API di basso livello più vicina allo state store (e quindi più lontana dall'astrazione utilizzata con la DSL).

Un esempio di una stream application che utilizza la stream DSL è dato dal seguente codice Scala<sup>2</sup>:

```
package io.confluent.examples.streams

import java.util.Properties

import org.apache.kafka.common.serialization._
import org.apache.kafka.streams._
import org.apache.kafka.streams.kstream.{KStream, KStreamBuilder}

object MapFunctionScalaExample {
  def main(args: Array[String]) {
    val builder = new KStreamBuilder

    val streamingConfig = {
      val settings = new Properties()
      settings.put(StreamsConfig.BOOTSTRAP_SERVERS_CONFIG, "localhost:9092")
      settings.put(StreamsConfig.DEFAULT_KEY_SERDE_CLASS_CONFIG,
        Serdes.ByteArray.getClass.getName)
      settings.put(StreamsConfig.DEFAULT_VALUE_SERDE_CLASS_CONFIG,
        Serdes.String.getClass.getName)
      settings
    }

    val text: KStream[Array[Byte], String] = builder.stream("input-topic")
    val upperCasedText: KStream[Array[Byte], String] =
      textLines.mapValues(_.toUpperCase())
    upperCasedText.to("output-topic")
  }
}
```

Nell'esempio partendo dal topic **input-topic** viene creato uno stream di record con chiave di tipo **ByteArray** e value (o contenuto del messaggio) di tipo **String**. Ad ogni messaggio viene applicata la funzione **.toUpperCase()** tramite la funzione **.mapValues()** definita dalla DSL (e quindi nodo della topologia di questa applicazione), infine viene pubblicato un nuovo record con il nuovo messaggio (la stringa in maiuscolo) su un nuovo topic `output-topic`.

---

<sup>2</sup> **Confluent Examples**, <https://github.com/diegoicosta/kafka-confluent-examples/blob/master/kafka-streams/src/main/scala/io/confluent/examples/streams/MapFunctionScalaExample.scala>

## 7. Conclusioni

Nei capitoli precedenti ho presentato le idee alla base dell'utilizzo di Apache Kafka, la sua architettura e il pattern (Event Sourcing) più comunemente utilizzato per sviluppare soluzioni software con la piattaforma.

È quindi ora necessario spiegare *perché* Apache Kafka è una soluzione alternativa ai comuni processi di ETL e il contesto che ha portato al suo sviluppo.

Nell'ultima decade la tipologia e mole di dati ed i sistemi di analisi/gestione di questi dati si sono trovati ad affrontare degli importanti cambiamenti nell'ambito dei processi di ETL (Neha, ETL Is Dead, Long Live Streams: Real-Time Streams with Apache Kafka, 2017).

In passato si era abituati a sistemi di gestione/analisi dei dati formati da più database operazionali che venivano utilizzati per creare una pipeline di dati diretta verso una o più data warehouse (Fig. 23). Questo processo era visto come un processo da eseguire poche volte al giorno e inutilizzabile per analisi real-time di dati.

L'approccio all'utilizzo dei dati ricavati era inoltre *passivo* più che *reattivo* data la scarsa necessità (o interesse) di essere rapidi nel rispondere agli eventi generati dall'analisi dei dati.

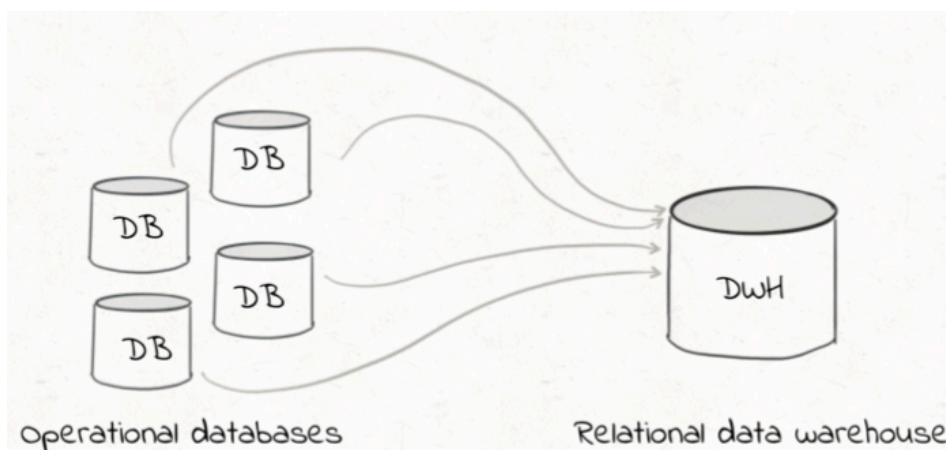


Figura 23: Database operazionali e data warehouse

Lo scenario che si è creato negli ultimi anni è però completamente diverso e può essere riassunto in pochi punti:

- Sempre più spesso, le vecchie architetture basate su un singolo database sono evitate in favore di architetture formate da microservizi distribuiti che raccolgono dati da più fonti
- Le tipologie di fonti di dati con cui interfacciarsi sono in rapida espansione, non esistono più solo database relazionali ma piuttosto, oltre agli stessi database relazionali, le aziende si trovano a dover fare uso di dati provenienti da database NoSQL, sensori, metriche e log delle tipologie più disparate e da un numero sempre crescente di sorgenti.
- La spinta verso lo *streaming* dei dati e una richiesta di aggiornamento dei dati "istantanea" porta il comune processo di ETL ad essere troppo lento per i moderni standard di analisi dei dati.

Questo non implica che un comune processo di ETL non può funzionare o che non sia possibile sviluppare delle soluzioni ad-hoc a seconda dei casi, il vero problema è che non avendo uno *standard*, una piattaforma o tecnologia con precise funzionalità atte a risolvere i problemi di questo scenario spesso ci si ritrova davanti a sistemi *over-engineered* o troppo complessi con capacità di sviluppo modulare basse e difficilmente integrabili con nuove tecnologie emergenti (Fig. 24).

Il risultato è che spesso una soluzione ad-hoc basata su un processo di ETL finisce per essere un groviglio di pipeline tra applicazioni, database e datawarehouse difficile da utilizzare e analizzare, soprattutto in caso di emergenze.

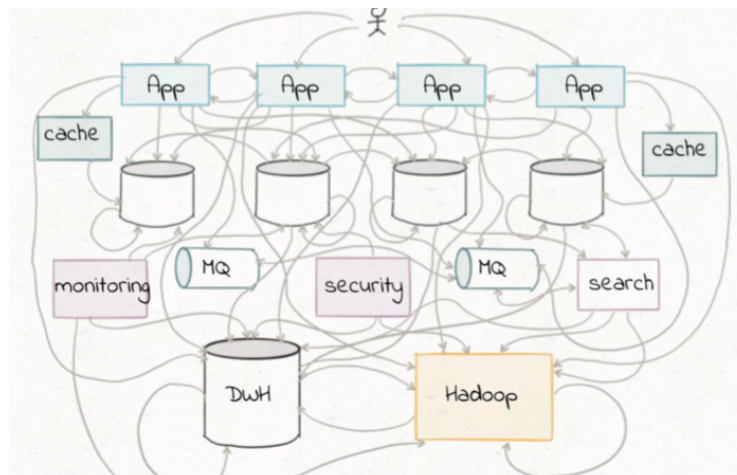


Figura 24: Soluzioni ETL ad-hoc

La necessità di sviluppare soluzioni ETL nasce durante gli anni '90, con la montante richiesta da parte della grande distribuzione di poter analizzare le abitudini di acquisto online dei propri clienti. In genere questo processo era

formato da tre fasi: l'estrazione di dati da più database operazionali, la trasformazione di questi dati in modo che aderissero allo schema della data warehouse ed infine il caricamento dei dati nella data warehouse. Con l'avanzare del tempo e della ricerca in questo campo si è però giunti alla conclusione che la *data coverage* nelle datawarehouse è spesso molto bassa rispetto ai dati presenti nei database operazionali, questo problema nasce dalla difficoltà nel analizzare e trasformare moli di dati importanti provenienti da più sorgenti (spesso con schemi di dati diversi) in gruppi di dati omogeni rispetto allo schema delle data warehouse.

Lo scenario moderno ha bisogno di piattaforme di real-time ETL, delle *streaming platform*, capaci di processare grossi volumi di dati delle tipologie più disparate, in tempo reale e capaci di adattarsi a possibili sviluppi futuri nella propria architettura (*forward-compatible*) come l'aggiunta di nuovi microservizi o l'utilizzo di nuove sorgenti di dati (Fig. 25).

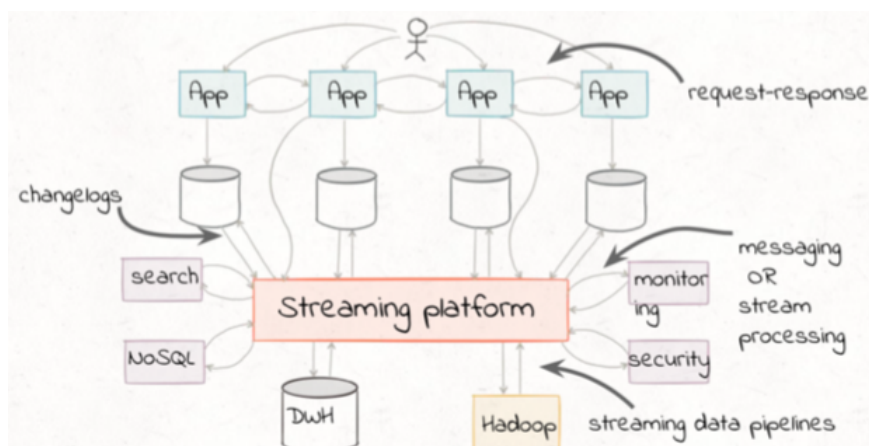


Figura 25: Architettura basata su streaming platform

Sono esattamente queste necessità che hanno portato un gruppo di sviluppatori di LinkedIn a sviluppare Apache Kafka: una streaming platform ormai utilizzata da moltissime grandi aziende nel ambito dei servizi web (*Netflix*, *PayPal* per citarne alcune). Kafka risolve tutti i problemi presentati precedentemente:

- La necessità di gestire grossi volumi di dati in tempo reale è risolta grazie alla sua architettura basata sul concetto di log, l'utilizzo dei producers, consumers e lo schema registry presentati nei capitoli precedenti.
- L'API Kafka Connect svolge il ruolo delle fasi di **E**xtraction e **L**oad dei processi di ETL, fornendo uno strumento capace di mettere in collegamento la piattaforma Kafka con una moltitudine di fonti di dati sia per la lettura che la scrittura delle trasformazioni.
- Il ruolo della fase di **T**ransform nei processi ETL è svolto dall'API Kafka Streams, la quale permette di sviluppare soluzioni software con il quale trattare e trasformare i dati pubblicati sulla piattaforma Kafka.



Apache Kafka è una tecnologia importante in continua crescita ed in continua adozione da moltissime realtà internazionali nell'ambito IT (alcuni esempi sono *The New York Times*, *Zalando* e *Pinterest*<sup>3</sup>), non per questo non ha debolezze.

Proprio perché è una nuova tecnologia in continua evoluzione le capacità di una azienda di adattare i propri sistemi all'utilizzo di Kafka sono strettamente legate alle loro capacità di adozione e studio di nuove tecnologie e soluzioni architetturali.

Nonostante le difficoltà tecnologiche non ci sono però dubbi che i processi di ETL stiano subendo una forte spinta evolutiva e che le piattaforme di streaming come Kafka rappresentano una valida alternativa ed il futuro dei processi di analisi e gestione di grossi volumi di dati.

---

<sup>3</sup> **Kafka Streams use cases**, <https://kafka.apache.org/documentation/streams/>

## Bibliografia

- Kleppmann, M. (2016). *Making Sense of Stream Processing*. O'Reilly Media Inc.
- Kreps, J. (2015). *Why Avro for Kafka Data?* Tratto da <https://www.confluent.io/blog/avro-kafka-data/>
- Neha Narkhede, G. S. (2017). *Kafka: The Definitive Guide: Real-Time Data and Stream Processing at Scale*. O'Reilly Media, Inc.
- Neha, N. (2016). *Event sourcing, CQRS, stream processing and Apache Kafka: What's the connection?* Tratto da Event sourcing, CQRS, stream processing and Apache Kafka: What's the connection?: <https://www.confluent.io/blog/event-sourcing-cqrs-stream-processing-apache-kafka-whats-connection/>
- Neha, N. (2017). *ETL Is Dead, Long Live Streams: Real-Time Streams with Apache Kafka*. Tratto da [www.youtube.com/watch?v=I32hmY4diFY](http://www.youtube.com/watch?v=I32hmY4diFY)
- Shapira, G. (2015). *Yes, Virginia, You Really Do Need a Schema Registry*. Tratto da <https://www.confluent.io/blog/schema-registry-kafka-stream-processing-yes-virginia-you-really-need-one/>
- Stopford, B. (2017). *Building a Microservices Ecosystem with Kafka Streams and KSQL*. Tratto da <https://www.confluent.io/blog/building-a-microservices-ecosystem-with-kafka-streams-and-ksql/>
- Stopford, B. (2017). *Messaging as the Single Source of Truth*. Tratto da <https://www.confluent.io/blog/messaging-single-source-truth/>
- Stopford, B. (2017). *Using Apache Kafka as a Scalable, Event-Driven Backbone for Service Architectures*. Tratto da <https://www.confluent.io/blog/apache-kafka-for-service-architectures/>
- Stopford, B. (2017). *Using Apache Kafka as a Scalable, Event-Driven Backbone for Service Architectures*. Tratto da <https://www.confluent.io/blog/apache-kafka-for-service-architectures/>
- Stopford, B. (2018). *Designing Event-Driven Systems*. O'Reilly Media Inc.