

Requisiti Software per Kafka e data streams per Digital Twins

Autore: Daniele Greco, Matias Negro, Nicolò Zaffaroni
Standard: ISO/IEC/IEEE 29148

Data: 27/11/2024
Stato: Bozza
Versione: 1.5

Sommario

1. Introduzione	1
1.1 Scopo del Documento	1
1.2 Scopo del Progetto	1
1.3 Panoramica	1
2. Descrizione Generale.....	2
2.1 Prospettiva del Prodotto	2
2.2 Funzionalità del Prodotto	3
2.3 Utenti e Utilizzo	3
2.4 Vincoli.....	3
2.5 Requisiti degli Stakeholder	3
3. Requisiti Specifici.....	4
3.1 Requisiti Funzionali	4
3.1.1 Tabella dei Requisiti Principali	4
3.1.2 Requisiti di Sistema.....	5
3.3 Requisiti di Sicurezza	6
3.4 Tracciabilità dei Requisiti	6
4. Specifiche Tecniche	7
4.1 Broker Kafka.....	7
4.2 Formato dei Messaggi	8
4.3 Containerizzazione	8
5. Appendici	8
5.1 Glossario	8
5.2 Dataset ed Esempi di messaggi	9
5.3 Riferimenti	10

1. Introduzione

1.1 Scopo del Documento

Questo documento specifica i requisiti per la progettazione e implementazione di una pipeline software per gestire lo scambio di dati tra il Physical Twin (simulazioni e sensori del paziente e del ventilatore) e il Digital Twin.

Il sistema è progettato per supportare:

- Trasmissione unidirezionale e bidirezionale di dati.
- Registrazione accurata dei messaggi in uno storage esterno.
- Scalabilità e modularità attraverso l'uso di container Docker.

1.2 Scopo del Progetto

L'obiettivo del progetto è implementare una pipeline dati che permetta:

1. La trasmissione unidirezionale e bidirezionale tra Physical e Digital Twin.
2. La registrazione completa di ogni scambio di dati.
3. L'archiviazione dei dati raccolti per analisi successive.
4. Controller REST per la gestione dei dati tra il Digital Twin e il Physical Twin, con un'implementazione che garantisca l'integrazione fluida con il sistema esistente. Il controller accetta richieste JSON per la configurazione del ventilatore e il monitoraggio del paziente, supportando trasmissioni sincrone con un tempo di risposta inferiore ai 100 ms.

1.3 Panoramica

Il documento include:

- Descrizione delle funzionalità richieste per la pipeline dati.
- Requisiti funzionali, non funzionali e di sicurezza.
- Specifiche tecniche per implementare la pipeline.

2. Descrizione Generale

2.1 Prospettiva del Prodotto

La pipeline funge da intermediario per la trasmissione dei dati tra:

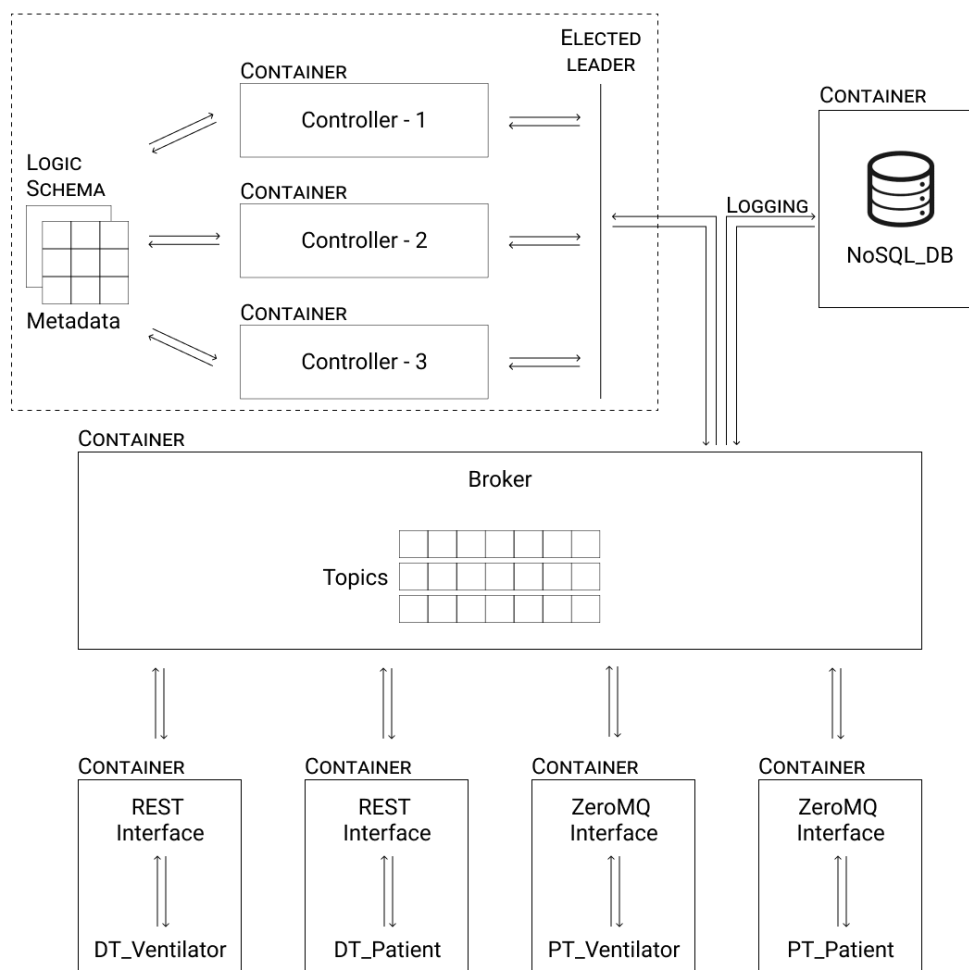
1. Paziente ospedaliero: Genera dati relativi allo stato di salute del paziente.
2. Ventilatore polmonare: Invio di parametri operativi e ricezione di istruzioni.
3. Digital Twin: Modello digitale che, attraverso i dati a lui forniti, simula il comportamento del relativo Physical Twin.

Il sistema includerà un wrapper per consentire l'integrazione tra il middleware ZeroMQ e Kafka, facilitando la comunicazione bidirezionale tra i componenti. Il wrapper ZeroMQ-Kafka traduce i messaggi pubblicati da ZeroMQ in messaggi Kafka, garantendo interoperabilità, ed è progettato per mantenere i metadati originali (timestamp, topic, tipo di messaggio) durante la conversione.

Tutti i dati saranno scambiati in formato JSON, con il middleware che implementa il pattern pub-sub tramite ZeroMQ. I seguenti topic sono definiti per gestire lo scambio dei dati:

- PatientParameters: Dati relativi al simulatore del paziente (es. frequenza cardiaca, volume polmonare).
- VentilatorParameters: Dati relativi al ventilatore (es. pressione, flusso).

Rappresentazione dell'infrastruttura Kafka:



2.2 Funzionalità del Prodotto

Descrizione delle funzionalità principali offerte dal sistema, evidenziando come ogni componente contribuisce al funzionamento complessivo. Le funzionalità sono orientate alla gestione dei dati tra il Physical Twin e il Digital Twin.

1. Trasmissione Dati: Trasferimento in tempo reale dei dati tra generatori e Digital Twin.
2. Registrazione e Logging: Ogni scambio di dati deve essere registrato con dettagli temporali.
3. Archiviazione: Salvataggio di tutti i dati in uno storage esterno.
4. Scalabilità: Possibilità di aggiungere nuovi componenti senza modificare l'architettura.

[DIAGRAMMA UML DEI CASI D'USO] Mostra gli attori (Operatori Medici, Ricercatori) e le loro interazioni con il sistema.

2.3 Utenti e Utilizzo

Possibili utenti del sistema con i relativi utilizzi di esso:

- Operatori medici: Monitorano i dati in tempo reale.
- Ricercatori: Analizzano i dati storici per migliorare il sistema.

2.4 Vincoli

Descrizione dei vincoli tecnici e operativi del progetto, tra cui i formati dati supportati, i protocolli utilizzati, e i limiti architetturali. Tali vincoli definiscono le regole entro cui il sistema deve operare.

- Formato dati: JSON per la trasmissione e archiviazione.
- Protocolli: ZeroMQ per la comunicazione con pattern Pub-Sub; REST per l'interazione tra componenti Digital Twin.
- Tempi di latenza: Ogni trasferimento deve essere processato in meno di 500 ms.

2.5 Requisiti degli Stakeholder

Elenco dei requisiti identificati dalle parti interessate del progetto, come operatori medici, ricercatori e sviluppatori. Questi requisiti guidano la progettazione del sistema per soddisfare le esigenze specifiche di ciascun gruppo

- Gli operatori medici richiedono dati affidabili in tempo reale per monitorare le condizioni dei pazienti e regolare il ventilatore.
- I ricercatori necessitano di un sistema scalabile per analizzare i dati storici e validare i modelli predittivi.
- Gli sviluppatori richiedono un'infrastruttura modulare basata su standard aperti (JSON, ZeroMQ, Kafka) per facilitare l'integrazione e la manutenzione.

3. Requisiti Specifici

3.1 Requisiti Funzionali

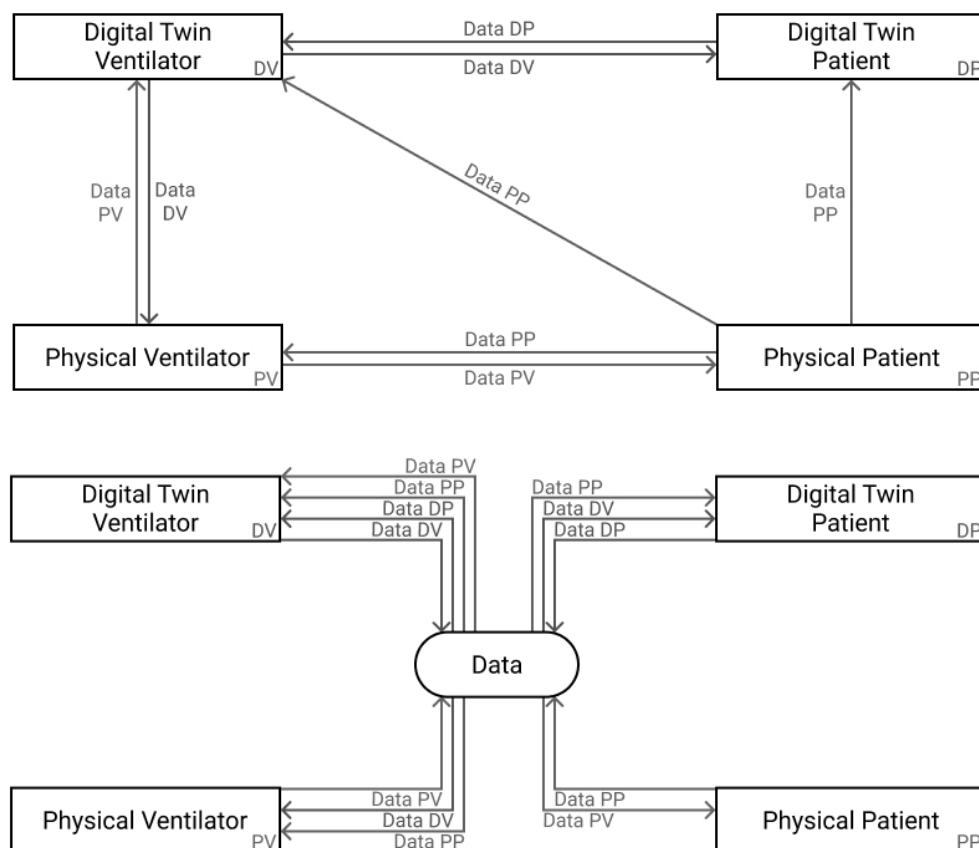
Descrizione dei requisiti funzionali e di sistema che il progetto deve soddisfare. La sezione include una tabella riepilogativa dei requisiti principali, una descrizione dei requisiti di sistema e una spiegazione dettagliata delle componenti chiave.

3.1.1 Tabella dei Requisiti Principali

La tabella seguente riassume i requisiti principali del sistema, organizzandoli per identificatore (ID) e descrizione. Rappresenta una panoramica dei requisiti chiave che saranno soddisfatti durante lo sviluppo.

ID	Requisito
RF-1	Il sistema deve inviare messaggi JSON dai simulatori al Digital Twin.
RF-2	Ogni messaggio deve includere timestamp, origine e tipo di dato.
RF-3	I messaggi ricevuti devono essere salvati in uno database NoSQL.
RF-4	Deve essere possibile aggiungere nuovi topic senza modificare l'architettura.
RF-5	Il Digital Twin deve essere in grado di ricevere e processare i dati in tempo reale dai simulatori e modificare le condizioni di ventilazione in base ai parametri ricevuti.

Diagrammi che illustrano lo scambio di dati tra le diverse entità della pipeline, inclusi i topic e i messaggi JSON:



3.1.2 Requisiti di Sistema

Descrizione dei requisiti di alto livello che il sistema deve soddisfare per garantire un'operatività ottimale.

- Il sistema deve supportare comunicazioni bidirezionali tra Physical Twin e Digital Twin.
- Deve essere possibile archiviare tutti i dati generati in uno storage esterno.
- Il sistema deve garantire la crittografia delle comunicazioni tramite SSL/TLS.
- Ogni componente del sistema deve essere containerizzato per garantire la portabilità.

3.1.3 Descrizione delle Componenti Chiave

Descrizione nei dettagli di ogni componente del sistema, con focus su protocollo, input/output e comportamento.

- Patient Simulator (Physical Twin Model)
 - Descrizione Generale: Il Patient Simulator è responsabile per la simulazione dei dati relativi al paziente, come la frequenza cardiaca, il volume polmonare e altri parametri vitali. I dati simulati sono essenziali per alimentare il Digital Twin del paziente e il ventilatore.
 - Protocollo: ZeroMQ
 - Input:
 - Formato: JSON
 - Unità di misura: Ogni dato deve includere l'unità di misura (ad esempio mL per volume, per_min per la frequenza respiratoria).
 - Frequenza nominale: Ogni 0.02 secondi.
 - Output:
 - Formato: JSON con struttura gerarchica che include il nome, il valore e l'unità di misura.
- Ventilator Software (Physical Twin)
 - Descrizione Generale: Il software del ventilatore fornisce i dati relativi al flusso d'aria e alle impostazioni del ventilatore, che influenzano direttamente i parametri del paziente.
 - Protocollo: ZeroMQ
 - Input: Stesso formato dell'output del Patient Simulator.
 - Output: Formato simile all'input del Patient Simulator, ma applicato ai parametri del ventilatore.
- ASM-Based Ventilator Model (Digital Model)
 - Descrizione Generale: Il modello ASM-Based Ventilator rappresenta la simulazione del ventilatore nel Digital Twin. Utilizza il software AsmetaS@Runtime per simulare il comportamento del ventilatore basato sui dati in ingresso.
 - Input: Lista separata da virgole con i nomi delle funzioni monitorate e i relativi valori.
 - Output: Lista separata da virgole con i nomi delle funzioni di output e i relativi valori.
 - REST-Based Server: Modalità di trasmissione sincrona (richiesta/risposta) per interazioni pull tra il middleware e il modello.

- **ASM-Based Digital Twin of the Patient**
 - Descrizione Generale: Il Digital Twin del Paziente è il modello virtuale che replica i dati fisiologici del paziente simulato, interagendo con il ventilatore per regolare i parametri vitali.
 - Input: Lista separata da virgole con i nomi delle funzioni monitorate e i relativi valori.
 - Output: Lista separata da virgole con i nomi delle funzioni di output e i relativi valori.
 - REST-Based Server: Modalità di trasmissione sincrona, con pull per entrambe le direzioni.

3.2 Requisiti Non Funzionali

I requisiti non funzionali descrivono le caratteristiche qualitative del sistema, come prestazioni, scalabilità e sicurezza. Questi requisiti completano quelli funzionali e definiscono i criteri per garantire un'esperienza utente ottimale.

ID	Requisito
RNF-1	La pipeline deve supportare almeno 10.000 messaggi al minuto.
RNF-2	Il sistema deve essere containerizzato per garantire portabilità.
RNF-3	Il sistema deve garantire <500 ms di latenza per i messaggi critici.
RNF-4	La latenza tra Pulse e BREATHE non deve superare 50 ms.
RNF-5	Il sistema deve permettere di scalare fino a 20 nodi Kafka senza perdita di messaggi.

3.3 Requisiti di Sicurezza

Definizione dei requisiti di protezione dei dati, autenticazione e crittografia richiesti per assicurare che il sistema sia conforme agli standard di sicurezza.

ID	Requisito
RS-1	Tutte le comunicazioni devono essere crittografate tramite SSL/TLS.
RS-2	Ogni componente deve essere autenticato con credenziali univoche.
RS-3	L'accesso ai dati deve essere regolato tramite un sistema Role-Based Access Control (RBAC), limitando l'accesso ai soli utenti autorizzati in base ai permessi assegnati.

3.4 Tracciabilità dei Requisiti

La matrice di tracciabilità dei requisiti collega i requisiti identificati agli stakeholder, alle implementazioni e ai metodi di verifica. Questa tabella garantisce che ogni requisito sia soddisfatto attraverso il design e test appropriati.

ID	Stakeholder	Implementazione	Verifica
RF-1	Operatori medici	Producer/Consumer con JSON su Pulse e BREATHE	Test di trasmissione dati real-time

RF-2	Ricercatori	Logger con timestamp JSON	Verifica integrità log
RF-3	Sviluppatori	Storage NoSQL per archiviazione	Test accesso e consistenza dati
RF-4	Sviluppatori, Operatori medici	Kafka con topic configurabili	Test aggiunta dinamica topic
RF-5	Operatori Medici, Ricercatori	Digital Twin processa i dati in tempo reale	Test latenza < 50 ms
RNF-1	Sviluppatori	Ottimizzazione Kafka per topic in Pulse	Test stress su topic
RNF-2	Sviluppatori	Containerizzazione con Docker	Verifica esecuzione in container
RNF-3	Sviluppatori, Operatori medici	Ottimizzazione ZeroMQ tra Pulse e BREATHE	Test latenza < 500 ms
RNF-4	Sviluppatori, Operatori medici	Scalabilità Kafka fino a 20 nodi	Test scalabilità e gestione failover
RS-1	Tutti	Implementazione SSL/TLS tra simulatori e Digital Twin	Test crittografia comunicazioni
RS-2	Tutti	Autenticazione con OAuth 2.0/JWT	Test autenticazione utenti
RS-3	Amministratori Sistema	Accesso ai dati basato su ruoli (RBAC)	Test permessi utente

4. Specifiche Tecniche

4.1 Broker Kafka

Descrizione dell'architettura e il ruolo del broker Kafka nella pipeline, spiegando come Kafka gestisce i topic, garantendo un flusso di dati affidabile tra le entità della pipeline.

Ogni messaggio inviato deve contenere un identificativo univoco per il paziente, basato sul nome del file di configurazione del paziente e una parte del percorso. Questa modifica garantisce la tracciabilità dei dati e la loro corretta gestione tra i vari componenti del sistema.

Topic	Descrizione
breathe-patient-parameters	Dati del paziente, inclusi parametri fisiologici e identificativo univoco (ID)
breathe-ventilator-parameters	Parametri operativi ventilatore.
breathe-simulation-output	Per i messaggi di regolazione inviati dal Digital Twin al Physical Twin.
breathe-feedback	Messaggi di regolazione inviati dal Digital Twin al Physical Twin.

pulse-action	Azioni applicate al paziente durante la simulazione (es. variazione parametri respiratori).
--------------	---

[DIAGRAMMA UML DI SEQUENZA] RAPPRESENTA IL FLUSSO DEI DATI DAL PATIENT SIMULATOR AL DIGITAL TWIN TRAMITE IL BROKER KAFKA

4.2 Formato dei Messaggi

Definizione del formato JSON utilizzato per lo scambio dei dati, essenziale per garantire una comunicazione coerente tra i componenti.

Campo	Descrizione
timestamp	Data e ora della trasmissione.
source	Identifica il modulo di origine.
type	Specifica il tipo di messaggio (input/output).
topic	Indica il topic su cui è stato inviato il messaggio.
status	Indica se la trasmissione è stata eseguita con successo.
payload	Contenuto del messaggio (JSON).

4.3 Containerizzazione

Ogni componente, incluso broker, producer e consumer, deve essere containerizzato. Ogni container sarà configurato con certificati SSL/TLS tramite un'autorità certificativa interna, dove le credenziali per l'autenticazione saranno gestite tramite un sistema OAuth2 centralizzato.

5. Appendici

5.1 Glossario

Physical Twin: Sistemi fisici o simulatori che generano dati, ovvero il paziente ospedaliero, il ventilatore polmonare.

PT: Abbreviazione di Physical Twin.

Digital Twin: Modelli digitali che replicano i comportamenti dei sistemi fisici.

DT: Abbreviazione di Digital Twin.

ZeroMQ: Middleware per la comunicazione peer-to-peer ad alte prestazioni.

Kafka: Broker di messaggistica distribuito per la gestione di topic.

ETL: Processo di estrazione, trasformazione e caricamento dei dati.

Broker: Sistema intermedio per la trasmissione dei messaggi (Kafka).

RF-‘numero’: ID relativo ad un requisito funzionale.

RNF-‘numero’: ID relativo ad un requisito non funzionale.

RS-‘numero’: ID relativo ad un requisito di sicurezza.

5.2 Dataset ed Esempi di messaggi

- Dataset:

I dataset saranno caricati nel Patient Simulator tramite script Python che convertono i file CSV in messaggi JSON.

- Respiratory and Heart Rate Dataset v1.0.0: Contiene dati fisiologici di pazienti con patologie respiratorie, utilizzati per addestrare e validare il modello BREATHE.
- ScienceDirect Respiratory Dataset: Dati reali utilizzati per modellare pazienti con ventilatori, con parametri di frequenza cardiaca e respiratoria.

- Esempi di Messaggi:

- breathe-patient-parameters:

Messaggio JSON contenente i parametri fisiologici del paziente simulato.

```
{
  "id": "patient_123_config_hash",
  "timestamp": "2024-02-20T14:25:43Z",
  "source": "breathe-patient-simulator",
  "type": "input",
  "payload": {
    "HeartRate": { "value": 75, "unit": "per_min" },
    "LungVolume": { "value": 1650, "unit": "mL" },
    "OxygenSaturation": { "value": 0.96, "unit": "%" }
  }
}
```

- breathe-ventilator-parameters:

Messaggio JSON contenente i parametri operativi del ventilatore.

```
{
  "id": "patient_123_config_hash",
  "timestamp": "2024-02-20T14:26:10Z",
  "source": "breathe-ventilator",
  "type": "input",
  "payload": {
    "VentilatorType": "Volume",
    "Pressure": { "value": 5, "unit": "mmHg" },
    "Volume": { "value": 20, "unit": "mL" }
  }
}
```

- breathe-simulation-output

Messaggio JSON di output della simulazione con parametri aggiornati del paziente.

```
{
  "id": "patient_123_config_hash",
  "timestamp": "2024-02-20T14:27:05Z",
```

```

"source": "breathe-simulator",
"type": "output",
"payload": {
  "SimTime": { "value": 14.579, "unit": "s" },
  "HeartRate": { "value": 72.0, "unit": "per_min" },
  "TotalLungVolume": { "value": 1652.87, "unit": "mL" },
  "RespirationRate": { "value": 12.0, "unit": "per_min" },
  "CarbonDioxide": { "value": 32.25, "unit": "mmHg" },
  "ArterialPressure": { "value": 103.43, "unit": "mmHg" }
}
}

```

- breathe-feedback

Messaggio JSON inviato dal Digital Twin per regolare i parametri del ventilatore.

```

{
  "id": "patient_123_config_hash",
  "timestamp": "2024-02-20T14:28:30Z",
  "source": "digital-twin",
  "type": "feedback",
  "payload": {
    "VentilationMode": "Assisted",
    "TargetPressure": { "value": 10, "unit": "mmHg" },
    "TargetVolume": { "value": 30, "unit": "mL" }
  }
}

```

5.3 Riferimenti

ISO/IEC/IEEE 29148:2018: *Systems and Software Engineering — Life Cycle Processes — Requirements Engineering*. International Organization for Standardization, Institute of Electrical and Electronics Engineers, 2018. Disponibile su IEEE Xplore: <https://ieeexplore.ieee.org/document/8559686>.

Kafka Documentation: *Apache Kafka Documentation*. The Apache Software Foundation, 2024. Disponibile su: <https://kafka.apache.org/documentation>

ZeroMQ Documentation: *ZeroMQ Guide*. The ZeroMQ Project, 2024. Disponibile su: <https://zeromq.org>.

Docker Documentation: *Best Practices for Containerized Applications*. Docker, 2024. Disponibile su: <https://docs.docker.com/>.

Goldberger AL, Amaral LAN, Glass L, et al.: *Respiratory Dataset v1.0.0*. PhysioNet, 2024. Disponibile su: <https://physionet.org/content/respiratory-dataset/1.0.0/>.

Bailón R, Sornmo L, Laguna P.: *Respiratory and Heart Rate Dataset*. ScienceDirect, 2024. Disponibile su: <https://www.sciencedirect.com/science/article/pii/S2352340923009460>.

Repository GitHub: *Pipeline per Digital Twin*. Repository GitHub, 2024. Disponibile su: <https://github.com/DaniGreco/Kafka-DigitalTwins>.

Repository GitHub: *Simulatore del corpo umano*. Repository GitHub, 2024. Disponibile su: <https://github.com/GionathaPirola/BREATHE>.