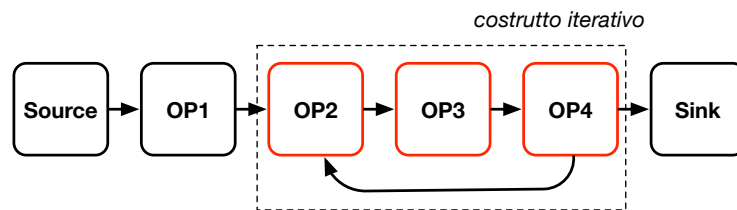


Titolo: Supporto di Costrutti DataFlow Ciclici per il Data Streaming in WindFlow

Proposta

Il paradigma del Data Stream Processing studia l'elaborazione efficiente di computazioni che operano su sequenze infinite di dati trasmessi ad alta velocità (es. dati finanziari, reti di sensori, social media). Le elaborazioni compiute possono andare dal calcolo di statistiche (streaming analytics) ma anche elaborazioni più complesse come il processo di inferenza su modelli precedentemente addestrati oppure tecniche di learning continuo. Dal punto di vista generale, le applicazioni di streaming vengono descritte come grafi dataflow, in cui i nodi rappresentano operatori eseguendo un calcolo stabilito dagli utenti sugli input ricevuti, e in grado di produrre output trasmessi ad altri operatori del grafo nella forma di risultati intermedi.



Lo sviluppo di sistemi di streaming riguarda anche la tipologia di operatori forniti agli sviluppatori ed incorporabili in grafi applicativi in grado di elaborare stream di input producendo output finali. Mentre solitamente i grafi dataflow sono aciclici, un controllato utilizzo di pattern ciclici può essere utile al fine di modellare computazioni su stream che aggiornano un modello mantenuto da un operatore sulla base di output prodotti da operatori successivi. Tale approccio è utile per modellare molti algoritmi e problemi tra cui quelli di Machine Learning.

Strumenti

La tesi è completamente incentrata sulla libreria open-source di streaming [WindFlow](https://github.com/ParaGroup/WindFlow) (<https://github.com/ParaGroup/WindFlow>), una libreria C++17 per lo streaming efficiente su multicore sviluppata dal Dipartimento di Informatica. L'obiettivo del lavoro prevede l'estensione del suo supporto a runtime per gli scopi definiti in questa proposta. Lo studente apprenderà quindi gli aspetti essenziali del linguaggio di programmazione C++ e dei suoi più recenti standard (C++17).

Obiettivo

Sistemi di streaming open-source in Java/Scala come Apache Flink supportano diverse famiglie di costrutti iterativi. La libreria WindFlow è attualmente deficitaria in questo senso, nonostante il supporto a runtime già contenga i meccanismi necessari per descrivere grafi con forme specifiche dotati di richiuse (e quindi di cicli). L'obiettivo del lavoro è quello di comprendere come modificare la API ad alto livello di WindFlow per consentire al programmatore di modellare grafi con cicli (rispettando opportuni vincoli), ed estendere il runtime della libreria per incorporare le modifiche necessarie al corretto funzionamento di questo pattern di utilizzo.

Prerequisiti

- Architetture degli Elaboratori e Sistemi Operativi
- Ingegneria del Software
- Una buona base di programmazione (con linguaggio C/C++)