

**Titolo:** Analisi del consumo energetico di strumenti di data stream processing su piattaforme Edge/IoT

## **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 eseguenti 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.

Le applicazioni di stream processing sono inerentemente parallele, essendo ogni nodo del grafo un thread indipendente. L'obiettivo di un buon motore di stream processing è quello di consentire l'esecuzione del grafo con la banda di elaborazione (throughput) più alta possibile e la latenza più bassa consentita. Le prestazioni si scontrano spesso con il consumo energetico, che dipende dalla frequenza di lavoro della CPU/GPU, dal numero di core attivi, e da altri aspetti legati all'architettura. L'obiettivo della tesi è quello di studiare il consumo energetico di sistemi di stream processing esistenti su dispositivi che mimano piattaforme di calcolo Edge, come le schede NVIDIA System-on-Chip (per esempio la Jetson Nano, ma anche più recenti soluzioni disponibili sul mercato).

## **Strumenti**

La tesi richiede di prendere dimestichezza con alcuni strumenti di data stream processing. In primo luogo Stream Processing Engine basati su JVM (Java Virtual Machine) come Apache Storm e Apache Flink. In secondo luogo, il linguaggio di programmazione C/C++ in riferimento alla libreria parallela per il data stream processing WindFlow, sviluppata dal Dipartimento di Informatica. Il candidato avrà a disposizione anche un pool di applicazioni di benchmark per i sopracitati strumenti, già disponibili in un repository pubblico GitHub.

## **Obiettivo**

Piattaforme NVIDIA System-on-Chip come la Jetson Nano (ma anche più recenti) forniscono un insieme di tool (es. `nvprof`) per configurare l'architettura dal punto di vista del consumo energetico consentendo di: disabilitare temporaneamente alcuni core della CPU; cambiando la frequenza di lavoro della CPU e della GPU; misurare il consumo energetico di un applicativo in esecuzione rispetto al consumo base della piattaforma. Tali strumenti possono essere utilizzati per capire l'impatto energetico di eseguire applicazioni di stream processing su queste piattaforme, comprendendone l'impatto dei parametri di configurazione e dei pattern paralleli di sviluppo, e derivandone così considerazioni quantitative e qualitative facendo, possibilmente, da apripista verso lo sviluppo di sistemi auto-configuranti che consentano a sistemi di stream processing di ottenere un buon bilanciamento tra prestazioni e consumo.

Il Candidato potrà quindi fare esperienza su questi strumenti e tecniche, partendo da un'analisi preliminare della letteratura per capire le modalità più accurate e robuste per l'analisi del consumo energetico di workload sulle piattaforme considerate nella tesi.

## **Prerequisiti**

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