





# Mind the Cost of Telemetry Data Analysis



Alessandra Fais<sup>1</sup>, Gianni Antichi<sup>2</sup>, Stefano Giordano<sup>1</sup>, Giuseppe Lettieri<sup>1</sup>, Gregorio Procissi<sup>1</sup>

<sup>1</sup> Università di Pisa, Italy | <sup>2</sup> Queen Mary University of London, United Kingdom

## Introduction

- ❖ Stream processing engines **efficiently process continuous amounts (streams) of information**
  - Widely used solutions (     ) for a variety of use cases
- ❖ Network operators need efficient ways to analyze fine-grained telemetry data
  - In production datacenter networks, hundreds of thousands of switches produce up to millions of reports per second!

**GOAL:** What's the best streaming engine for network traffic analysis?

## A Qualitative Comparison

	Flink	Spark	Storm	WindFlow
Batching	✗	mandatory	✗	✓
Chaining	✓	✗	✗	✓
Ordering	✗	btw batches	✗	✓
Windows	✓	✓	✗	✓
Event time	✓	partial	✗	✓
Distributed	✓	✓	✓	✗

## Findings

- Systems designed for generic data processing over distributed platforms perform poorly with network data
  - Overheads not compensated by the computational burden of the application itself

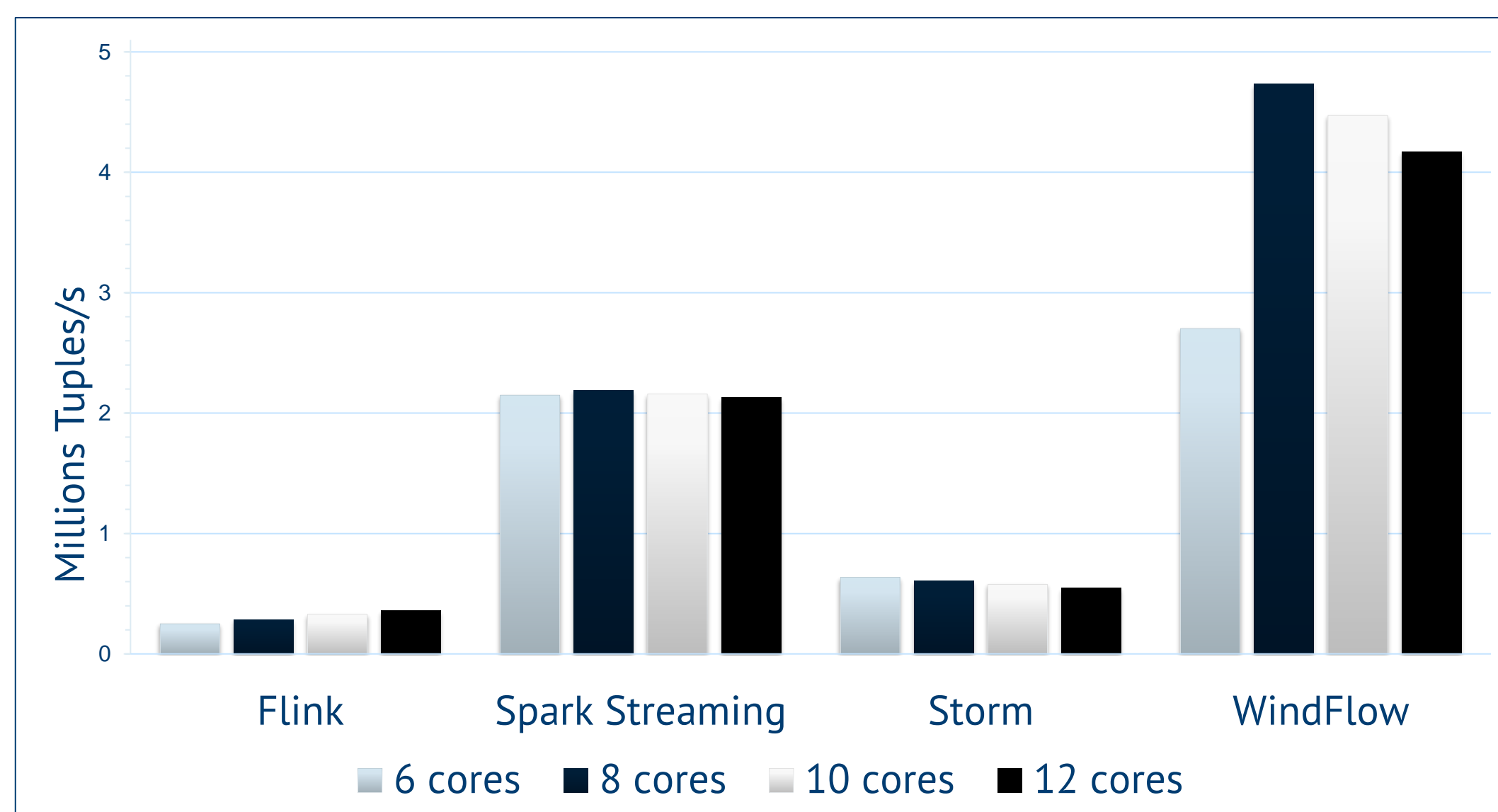
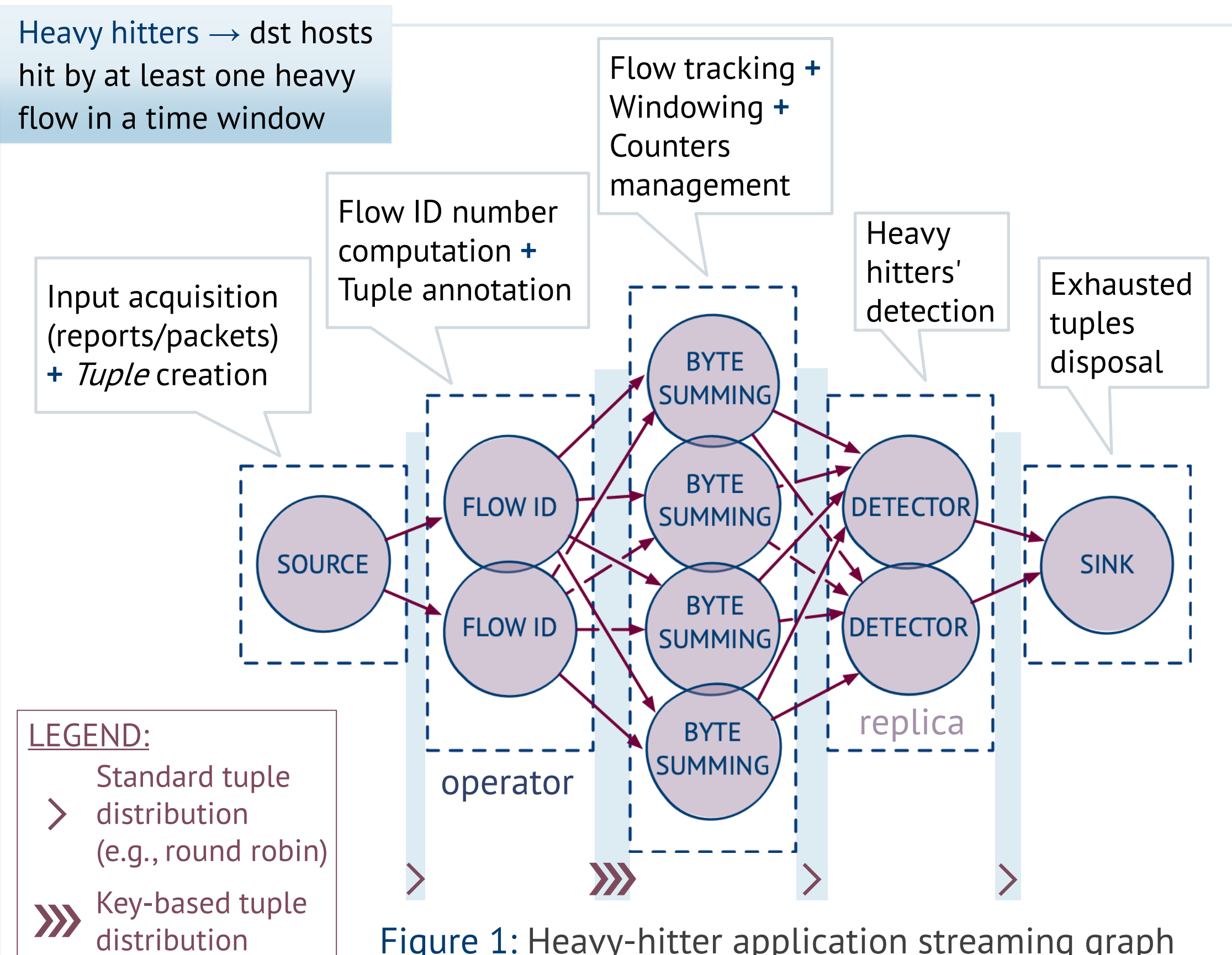


Figure 2: Performance comparison

- WindFlow shows better performance figures
  - More than 2x in most cases than Spark Streaming
  - Around 10x of Flink and Storm
- WindFlow performance scales with n. physical cores
- Other solutions immediately saturate

### Resource utilization scenarios

- Physical cores only → number of cores  $\leq 8$
- Hyperthreading →  $8 < \text{number of cores} \leq 16$

## Promising research directions

### Design of a networking domain specific streaming engine

