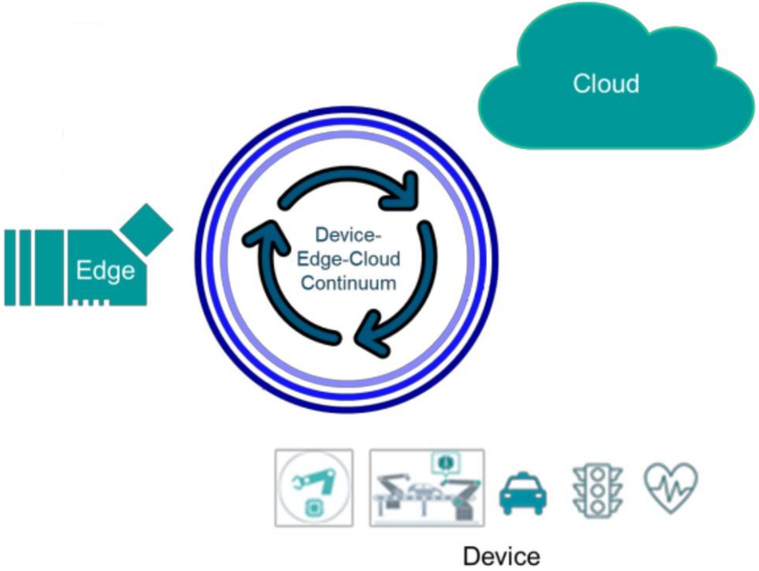
Semantic Stream Processing on different Hardwares.

Anh Le-Tuan

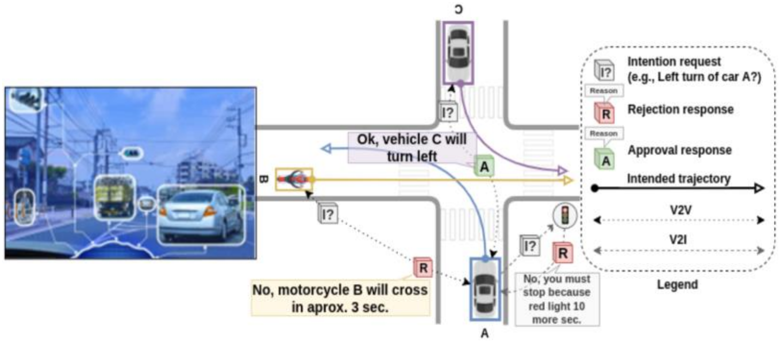
SMARTEDGE

- enable the **dynamic integration** of **decentralized intelligence** on the **cloud-edge continuum**.

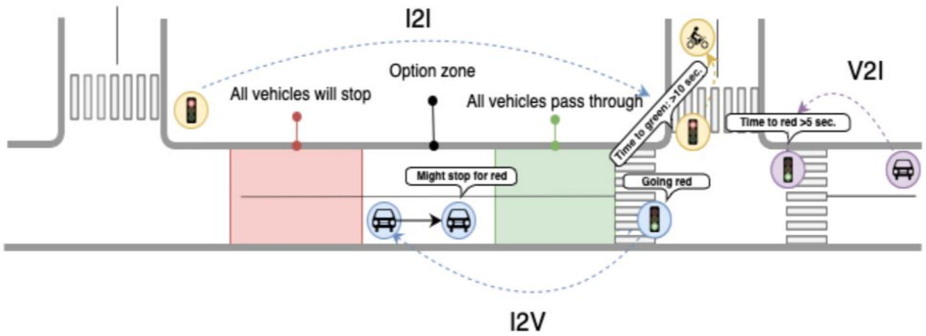
- ensure **robustness**, **reliability**, security, privacy and **scalability**.



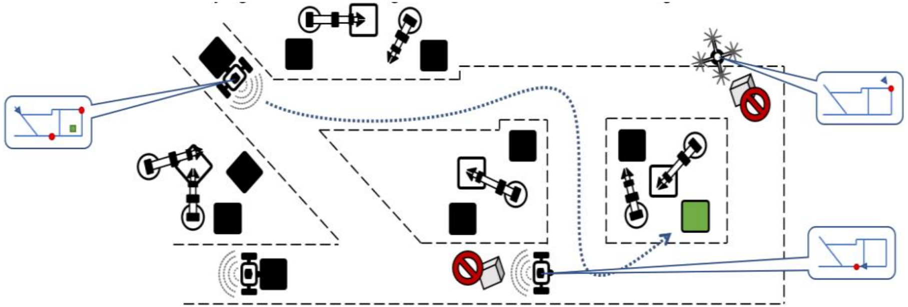
SMARTEDGE Scenarios



Scenario 1: Cooperative Perception for Driving Assist

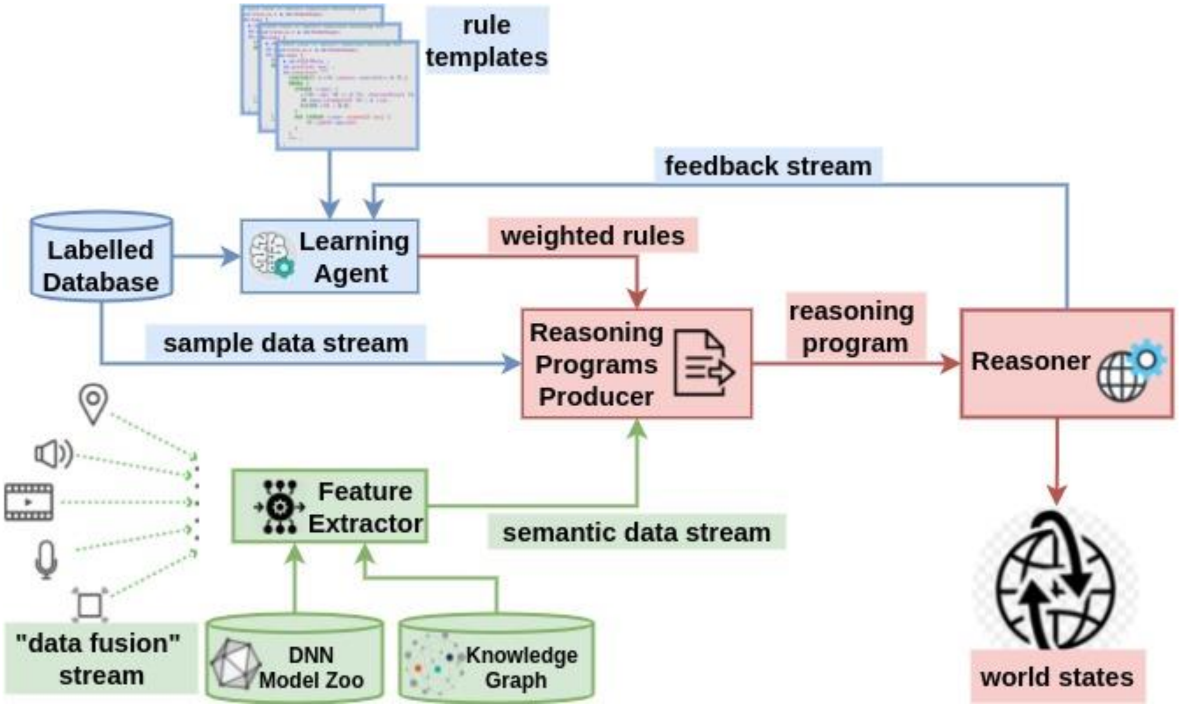


Scenario 2: I2V Intelligence Swarm



Scenario 3: Collaborative Robotic Movers

Semantic-driven Multimodal Stream Fusion Framework



The overview of conceptual design of CQELS 2.0

Enable **multimodal data stream fusion** via **probabilistic stream reasoning.**

Challenges

Write **once** run **everywhere** but :

- Scalable
- Low latency

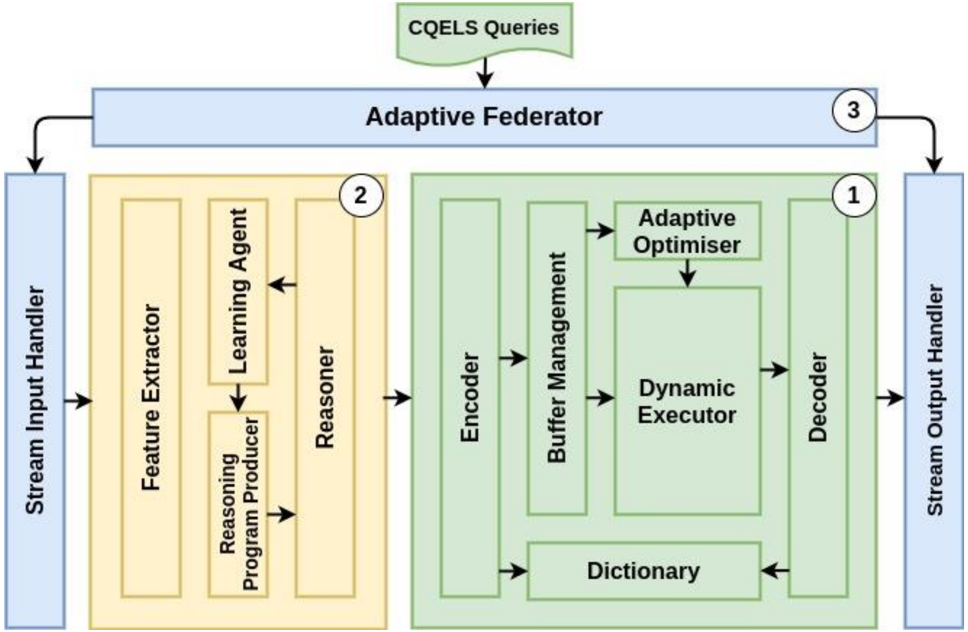
CQELS 2.0 Architecture

Three subsystems:

1. Semantic symbolic stream processor

2. Multimodal data stream feature extractor.

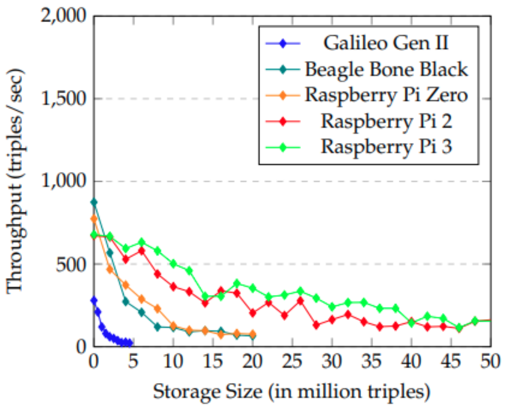
3. Adaptive federator.



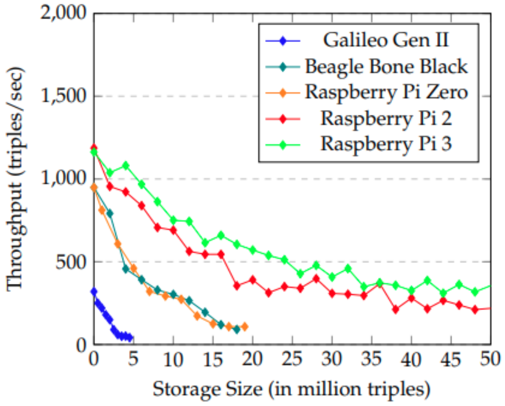
The overview of architecture of CQELS 2.0

Le-Tuan, A., Nguyen-Duc, M., Le, C. Q., Tran, T. K., Hauswirth, M., Eiter, T., & Le-Phuoc, D. (2022). CQELS 2.0: Towards A Unified Framework for Semantic Stream Fusion. *arXiv preprint arXiv:2202.13958*.

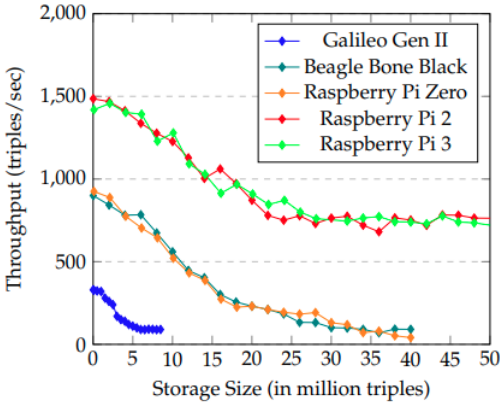
I/O Bottleneck for flash storage of Edge devices



(a)JenaTDB



(b)RDF4J



(c)Virtuoso

	Data Structure	Version
Jena TDB	B^+tree LRU Cache	3.14.0
RDF4J Native Store	BTree	3.1.0
Virtuoso Open-Source	B^+Tree	6.1.8

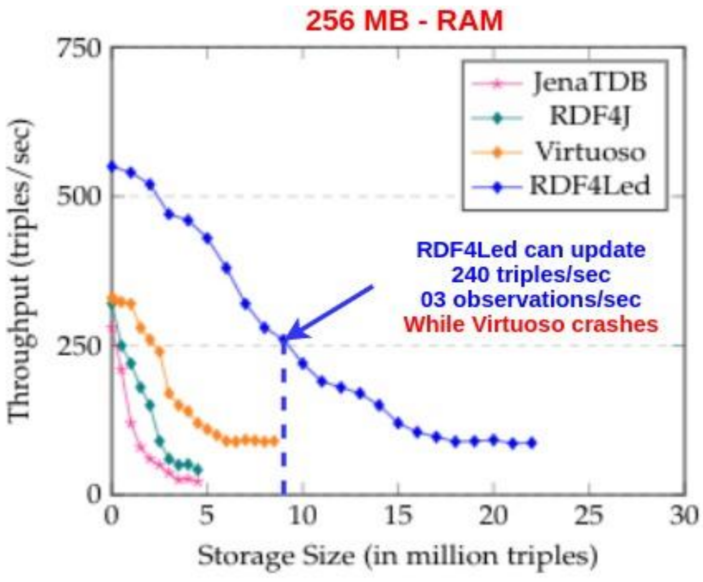
	RAM	CPU
Galileo II (GII)	256 MB	0.4 GHz
Raspbery Pi Zero (RPi0)	512 MB	1.0 GHz
Raspbery Pi Zero W (BBB)	512 MB	1.0 GHz
Raspbery Pi 2 B+ (RPi2)	1 GB	0.9 GHZ (4 cores)
Raspbery Pi 3 B+ (RPi3)	1 GB	1.2 GHZ (4 cores)

Get risk of I/O bottleneck

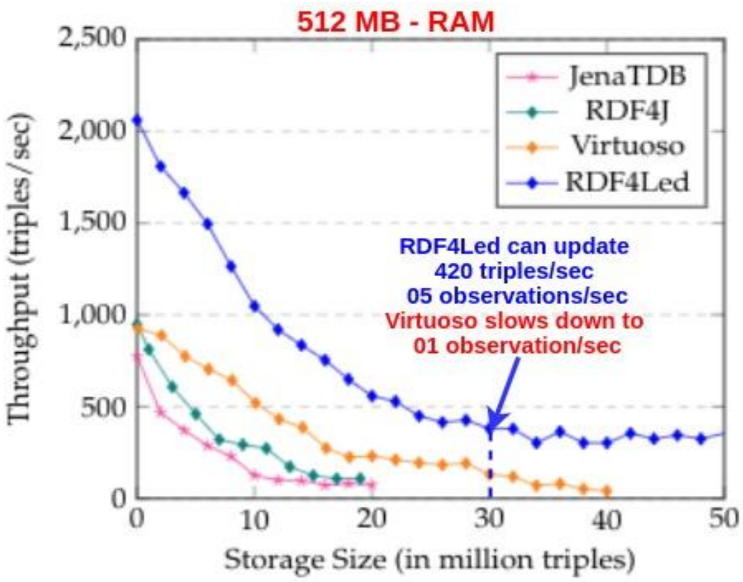
- Optimize memory footprint.
- Flash-friendly I/O data structure.
- Cold Clean First – RLU replacement policy.

Le-Tuan, A., Hayes, C., Hauswirth, M., & Le-Phuoc, D. (2020). Pushing the scalability of RDF engines on IoT edge devices. *Sensors*, *20*(10), 2788.

RDF4LED - Write throughput

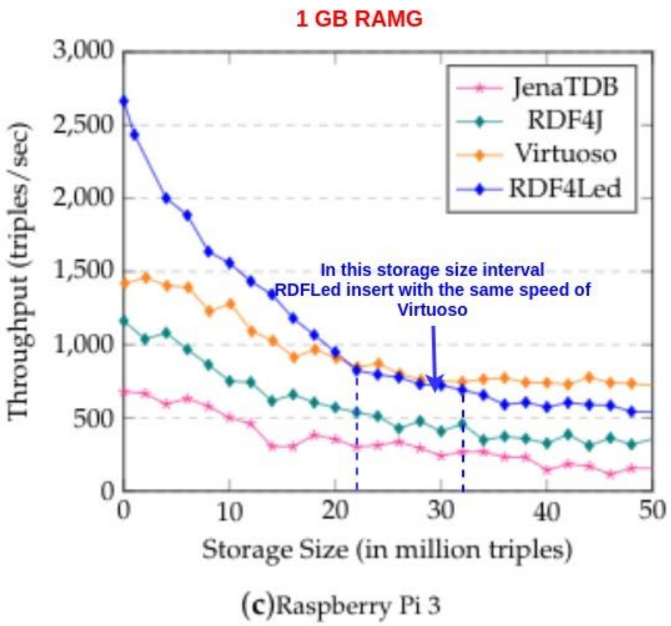


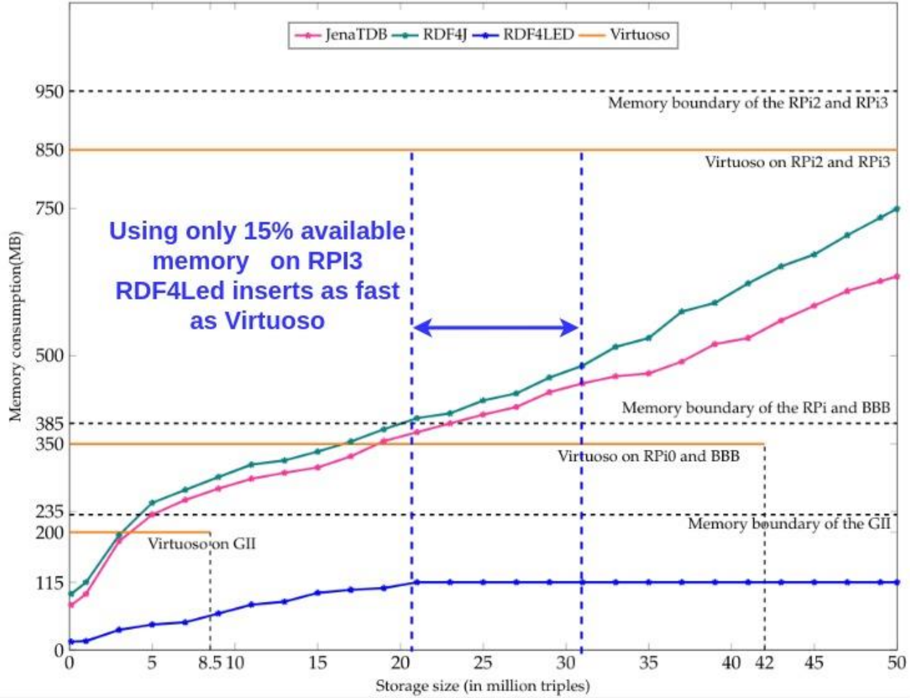
(a)Gallileo Gen II



(b)Raspberry Pi Zero;

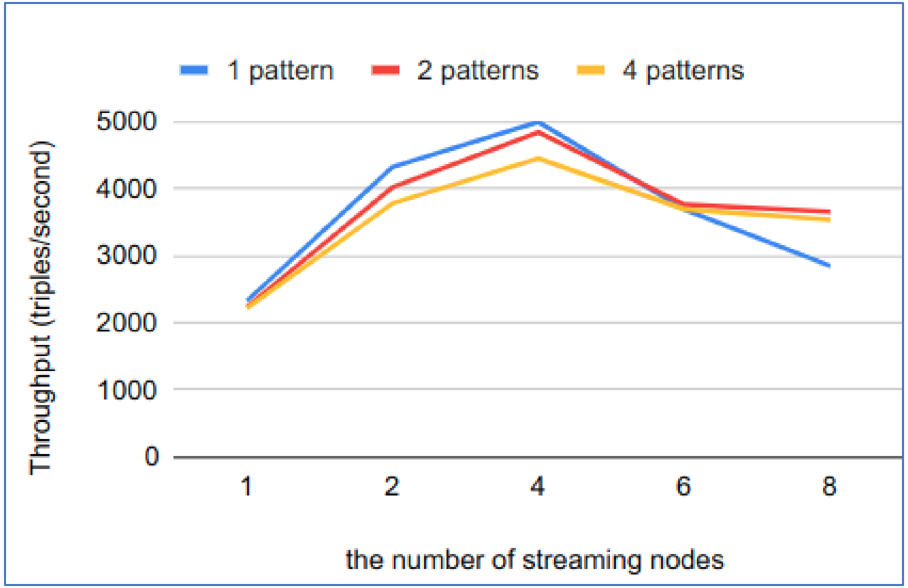
RDF4Led – Write throughput and memory consumption





Scaling out Issue

- Adding **more** nodes **increase** the through.
- Only adding **more** nodes does not.

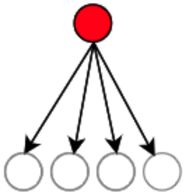


Nguyen-Duc, M., Le-Tuan, A., Calbimonte, J. P., Hauswirth, M., & Le-Phuoc, D. (2020). Autonomous RDF stream processing for IoT edge devices. In *Joint International Semantic Technology Conference.*

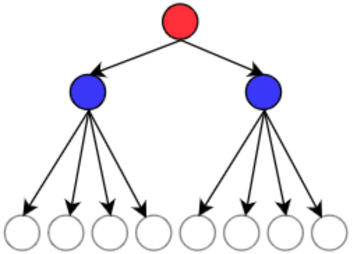
Scaling out Issue

- How to coordinate?
- Where to place processing operators.

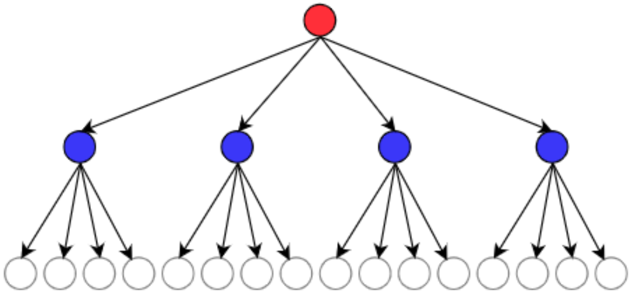
Setup topologies



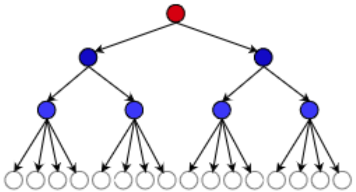
(a) 1 node (1-hop)



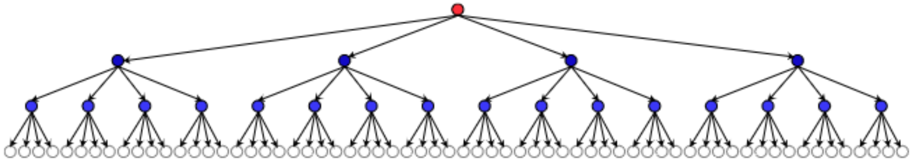
(b) 3 nodes (2 hop, 2-fanout)



(c) 5 nodes (2 hop, 4-fanout)

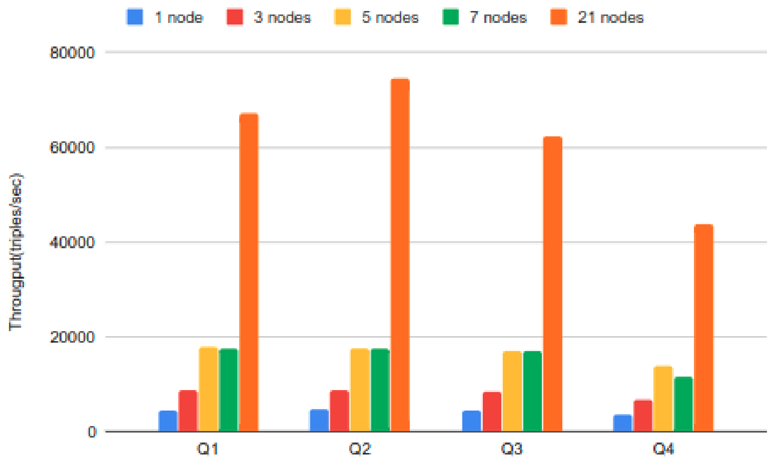


(d) 7 nodes (3-hop, 2-fanout)

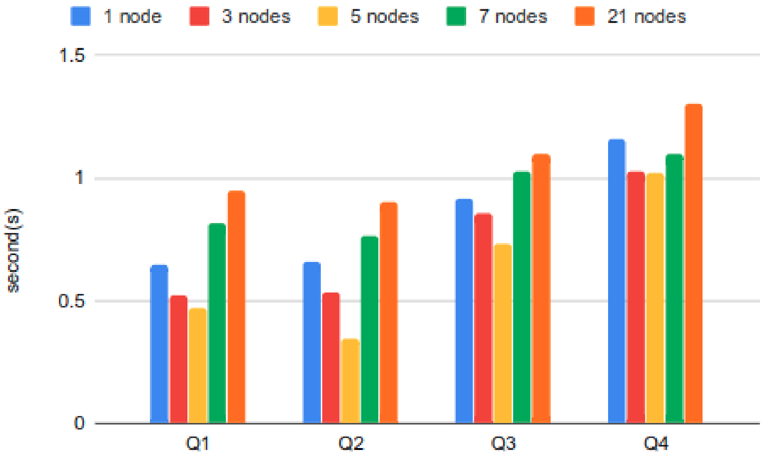


(e) 21 nodes (3-hop,4-fanout)

Distributed Stream results



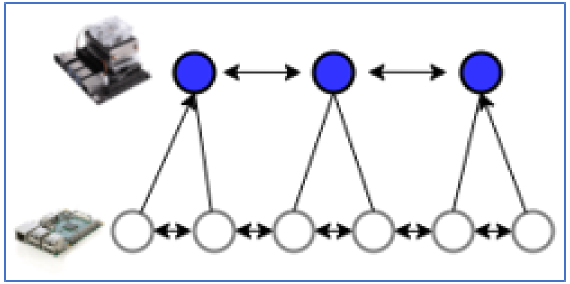
(a) Processing Throughput

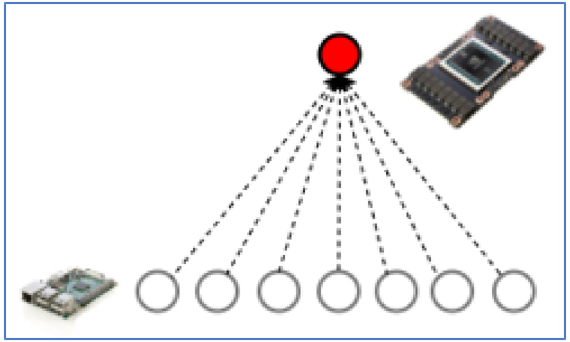


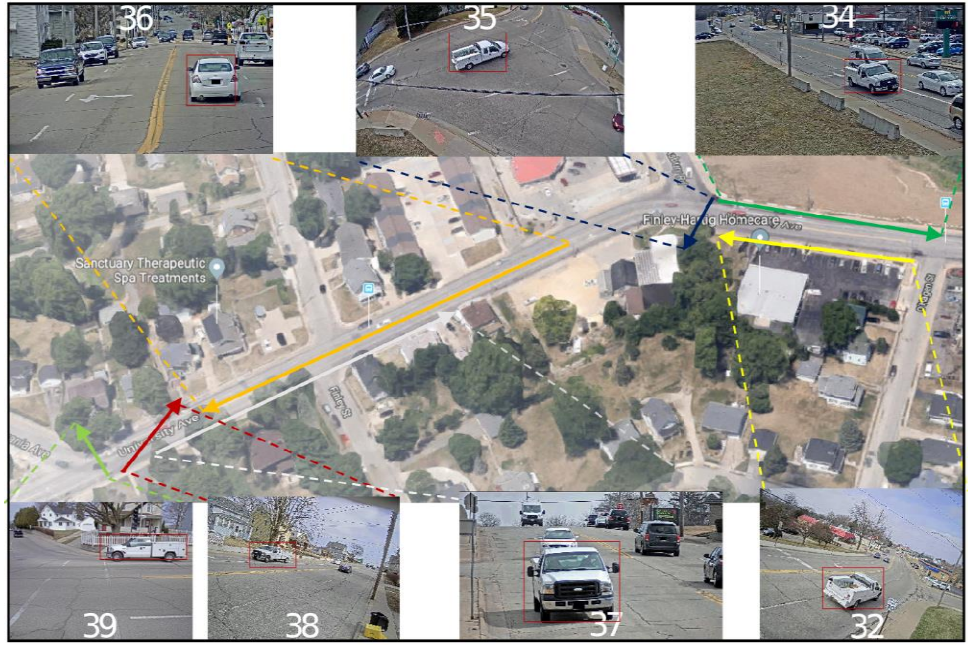
(b) Average Processing Time

Open Distributed Sytems
ODS

Offloading feature extraction



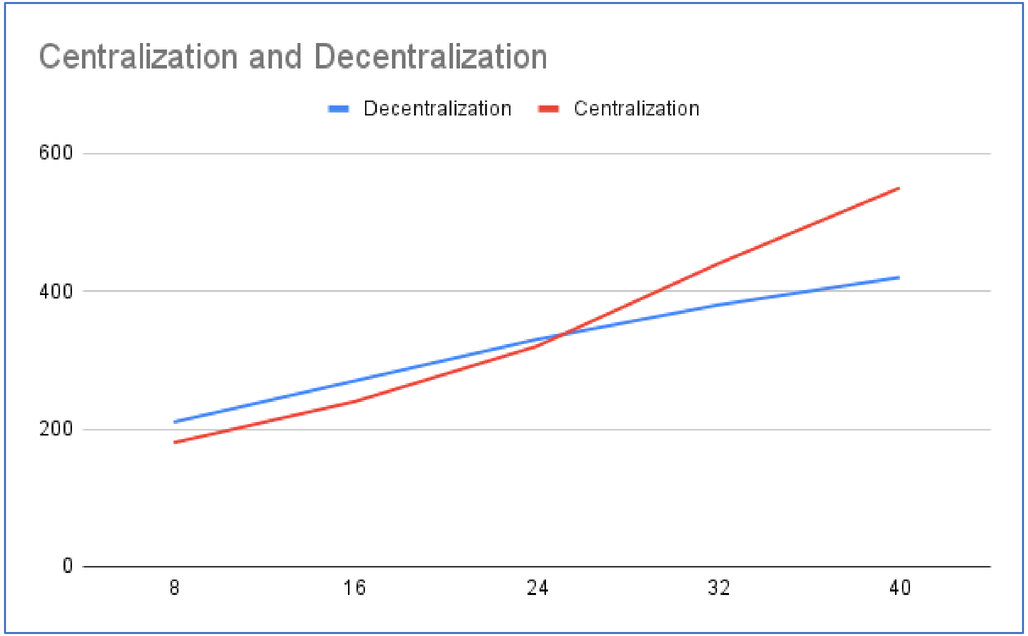




Nguyen-Duc, M., Le-Tuan, A., Hauswirth, M., & Le-Phuoc, D. (2021, June). Towards autonomous semantic stream fusion for distributed video streams. In *Proceedings of the 15th ACM International Conference on Distributed and Event-based Systems* (pp. 172-175).

Traffic camera edge network

	8 x Jetson Nano	2 x Server
RAM	8 x 4GB	2x 1TB RAM
CPU	8 x Quadcore ARM Cortex-A57	2x 10 physical cores
GPU	NVIDIA Maxwell 128 NVIDIA CUDA cores.	2 x NVIDIA Tesla V100 16GB
Inference	8x 15-17fps	2 x 250 fps



Open Issue

Accelerating Execute Engine with different factors:

- Faster indexing:

1. On multiple type of storage : DDDR2 RAM, GPU RAM, flash, SSD, HDD
2. On different processing hardwares: Intel, arm, snapdradon, Nvidia, ...

- Integrating ML or DNN components/models engine: Pytorch, DL2/D4JL, TensorFlow/TensorLight,

Open Issue

- Better Coordination:
    1. Communication: Networking, Topology, Bandwidth,
    2. Operator placement
    3. Resource management.