Presenting at

Karlsruher Institut für Technologie

2nd BMBF Big Data All Hands Meeting and
2nd Smart Data Innovation Conference
Karlsruhe , October 11.-12., 2017

Technische Universität Berlin

Efficiently Handling Streams from Millions of Sensors

Jonas Traub – TU Berlin / DFKI

The Growth of the Internet of Things

BY THE YEAR 2020, THERE WILL BE [IDC, Big Data in IoT, 2014]

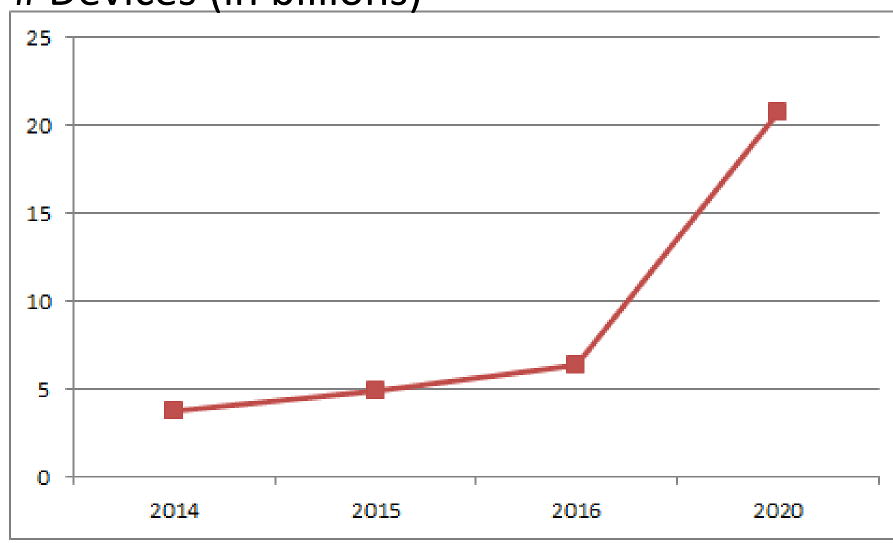
50,000,000,000 connected devices, creating and sharing

40,000,000,000,000 GB

worth of data across the Internet of Things.

Gartner says 6.4 billion connected "Things" will be in use in 2016 and more than 20 billion in 2020.

\# Devices (in billions)



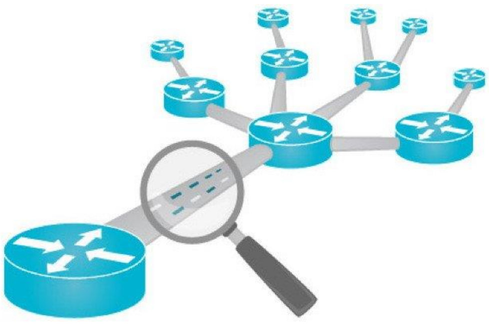
Year

Goal

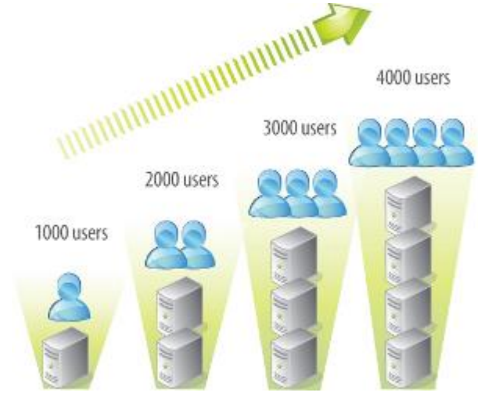
Provide real-time insights based on IoT data.

Problem

- Billions of devices provide real-time data
- Result: Vast amount of data streams



Heavy Network Utilization



Scalability Challenges

Increasing Latencies

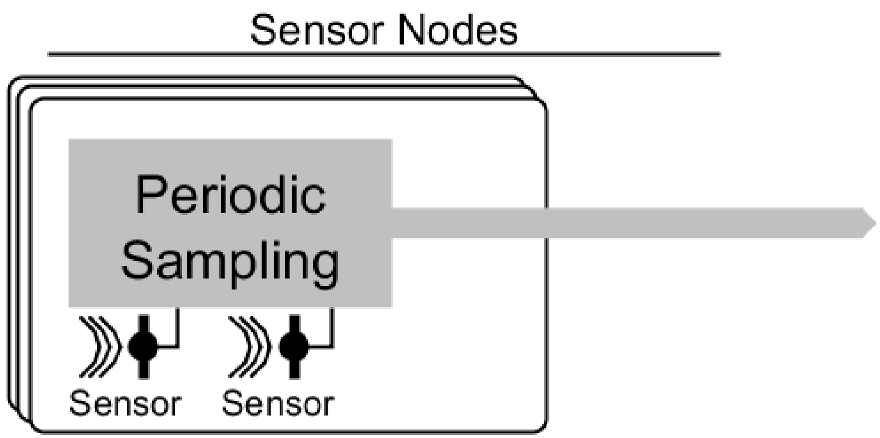
Financial Costs

Solution

Produce and process data streams based on the data demand of applications.

State of the Art Approach

Data Stream Production with Periodic Sampling

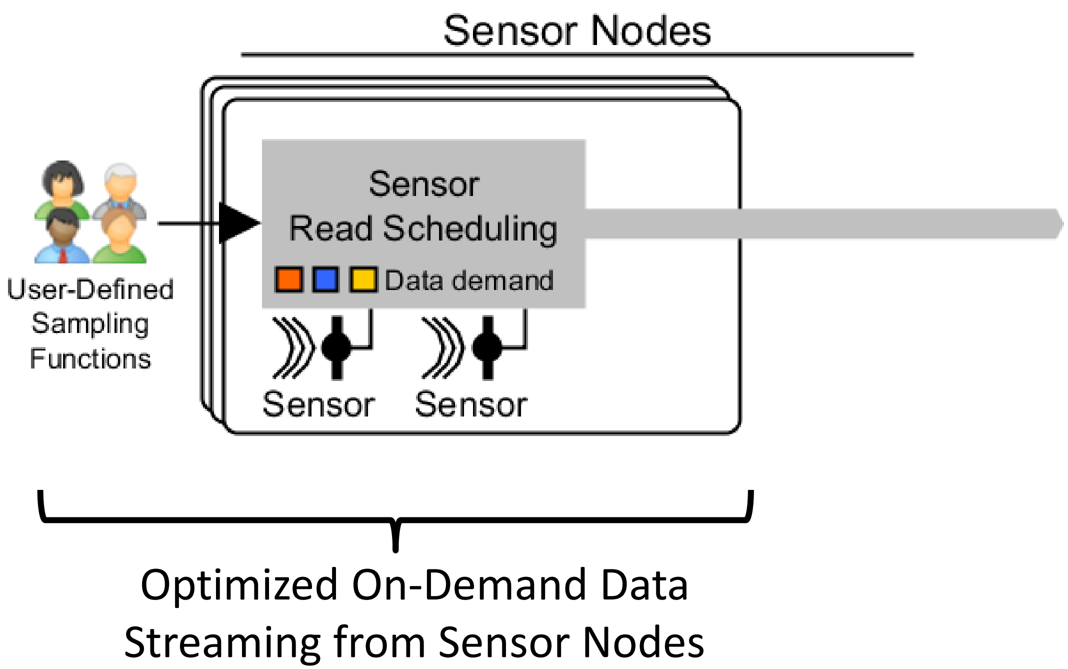


Major Challenges:

- Oversampling
- Missing Adaptivity

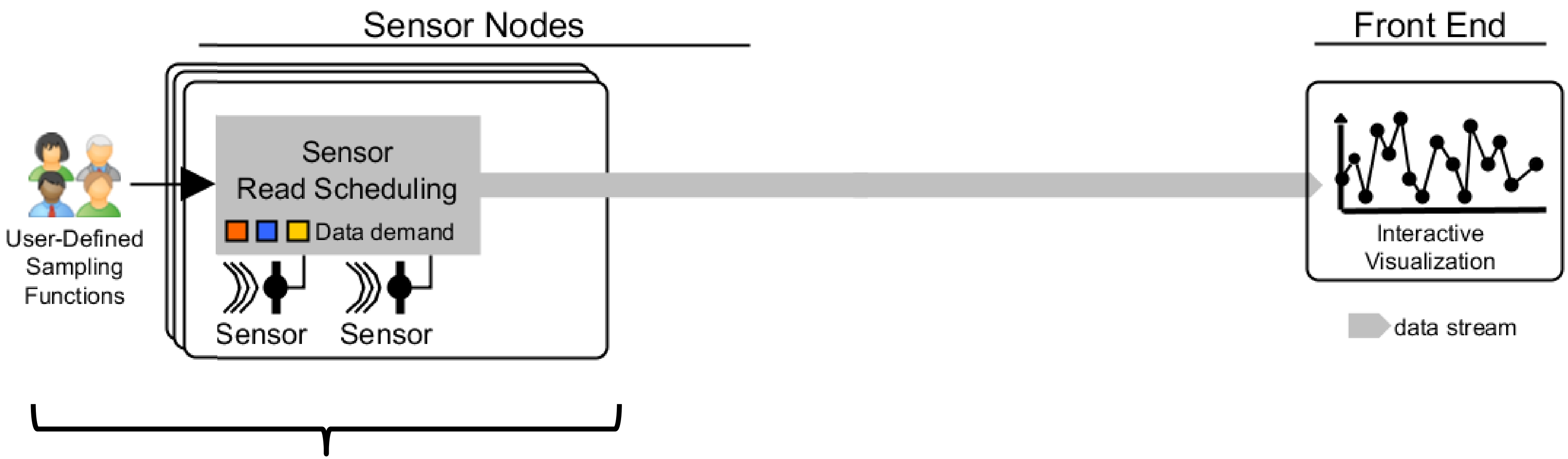
Solution

On-Demand Data Streaming from Sensor Nodes



State of the Art Approach

Provide all Data to Front-End Applications



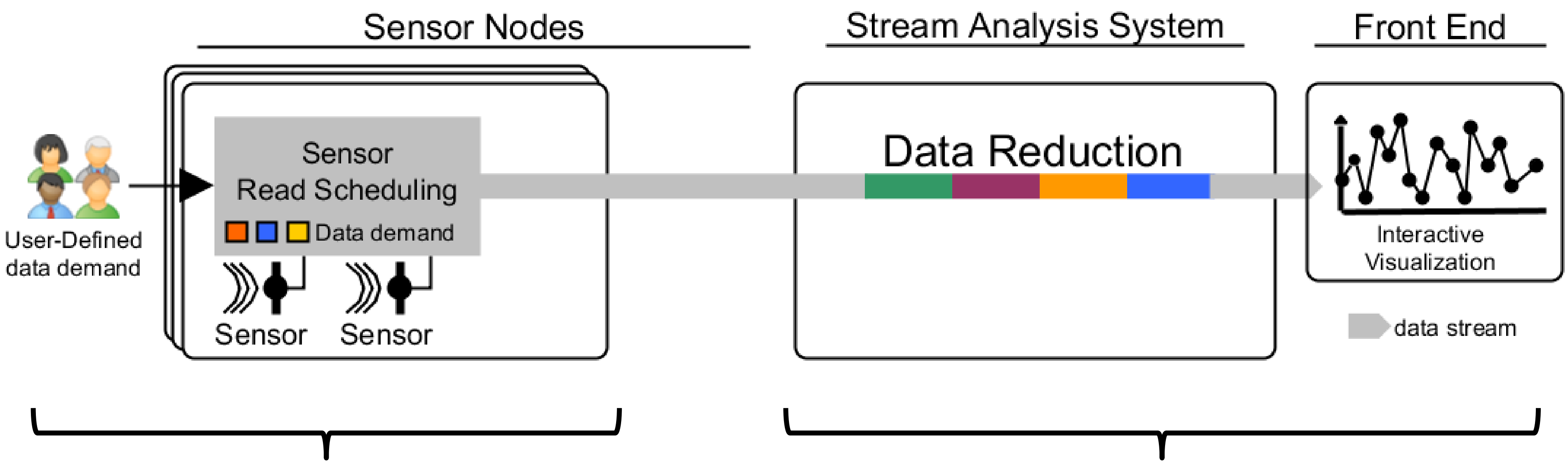
Optimized On-Demand Data Streaming from Sensor Nodes

Major Challenge:

- Front End Overload

Solution

Adaptive Data Reduction with Streaming Engines

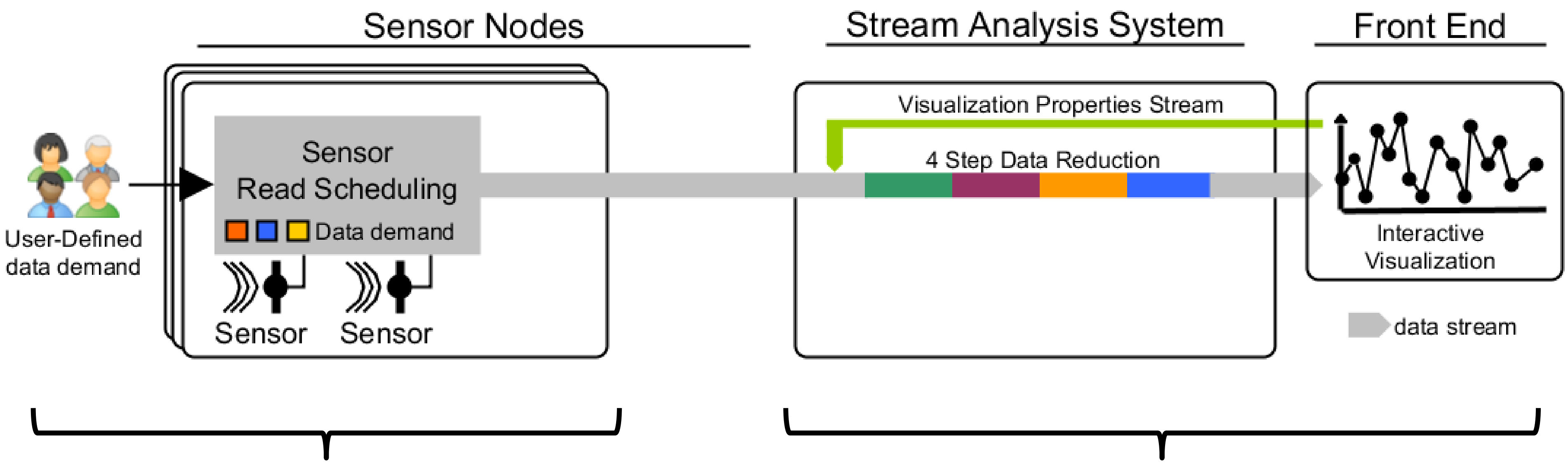


Optimized On-Demand Data Streaming from Sensor Nodes

I²: Interactive Real-Time Visualization for Streaming Data

Solution

Adaptive Data Reduction with Streaming

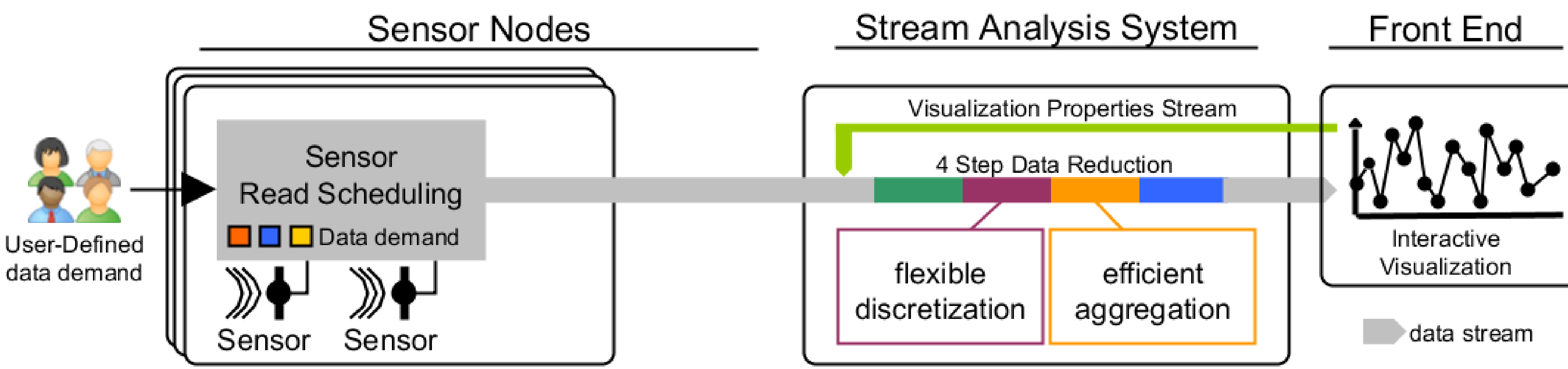


Optimized On-Demand Data Streaming from Sensor Nodes

I²: Interactive Real-Time Visualization for Streaming Data

Solution

Efficient Processing of user-defined Windows



Optimized On-Demand Data Streaming from Sensor Nodes

I²: Interactive Real-Time Visualization for Streaming Data

Cutty: Aggregate Sharing for User-Defined Windows

Publications

Optimized On-Demand Data Streaming from Sensor Nodes

Jonas Traub[1] Sebastian Breß[1,2] Tilmann Rabl[1,2] Asterios Katsifodimos[3] Volker Markl[1,2]

[1]*Technische Universität Berlin* [2]*German Research Center for Artificial Intelligence (DFKI)* [3]*SAP Innovation Center*

jonas.traub@tu-berlin.de sebastian.bress@dfki.de rabl@tu-berlin.de

Cutty: Aggregate Sharing for User-Defined Windows

Paris Carbone† Jonas Traub‡ Asterios Katsifodimos‡
Seif Haridi† Volker Markl‡

†KTH Royal Institute of Technology
{parisc,haridi}@kth.se

‡ Technische Universität Berlin & DFKI

I²: Interactive Real-Time Visualization for Streaming Data

Jonas Traub
Technische Universität Berlin
jonas.traub@tu-berlin.de

Nikolaas Steenbergen
German Research Center for Artificial Intelligence (DFKI)
nikolaas.steenbergen@dfki.de

Philipp M. Grulich
German Research Center for Artificial Intelligence (DFKI)
philipp.grulich@dfki.de

Tilmann Rabl
Technische Universität Berlin
rabl@tu-berlin.de

Volker Markl
Technische

ACM Symposium on Cloud Computing
Santa Clara, California, September 25-27, 2017

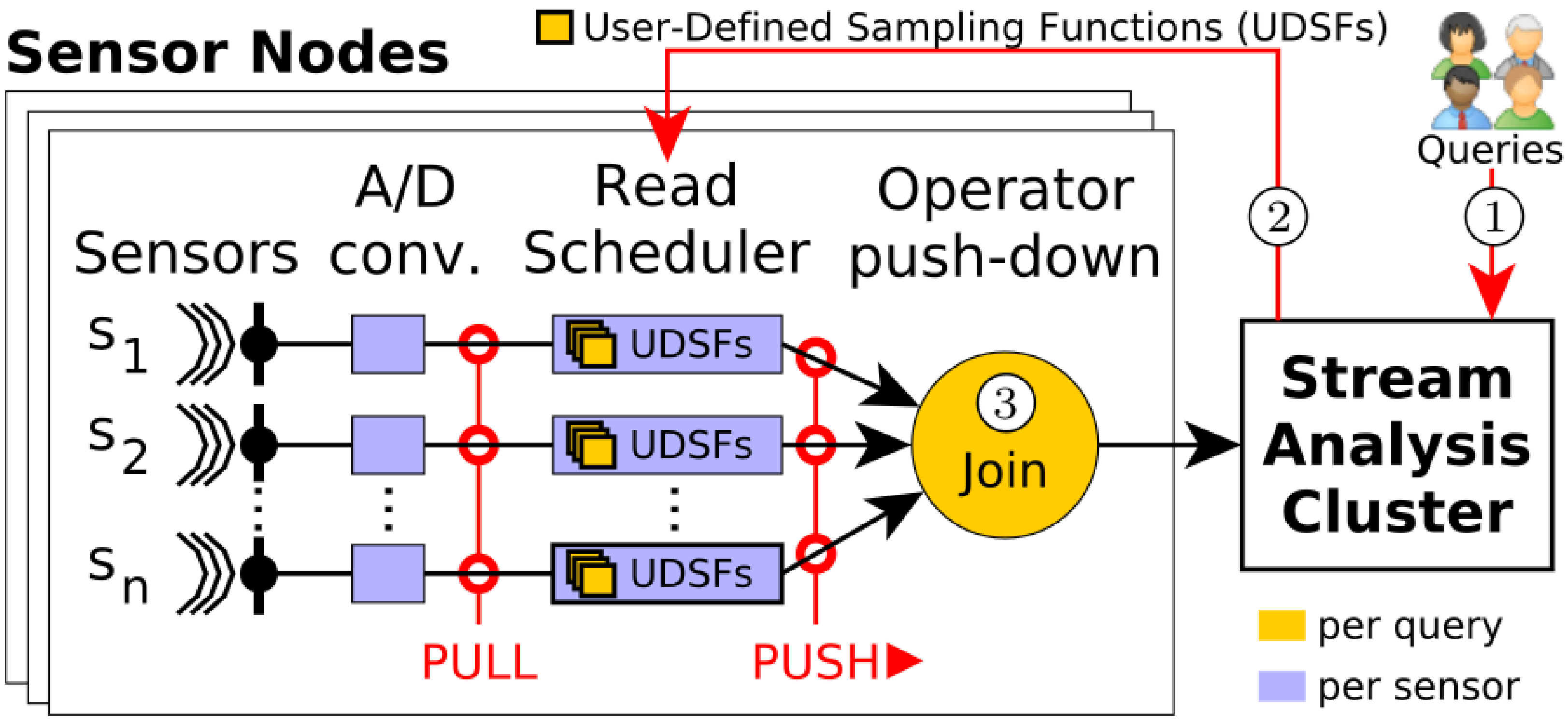
Optimized On-Demand Data Streaming from Sensor Nodes

Jonas Traub, Sebastian Breß, Asterios Katsifodimos, Tilmann Rabl, Volker Markl

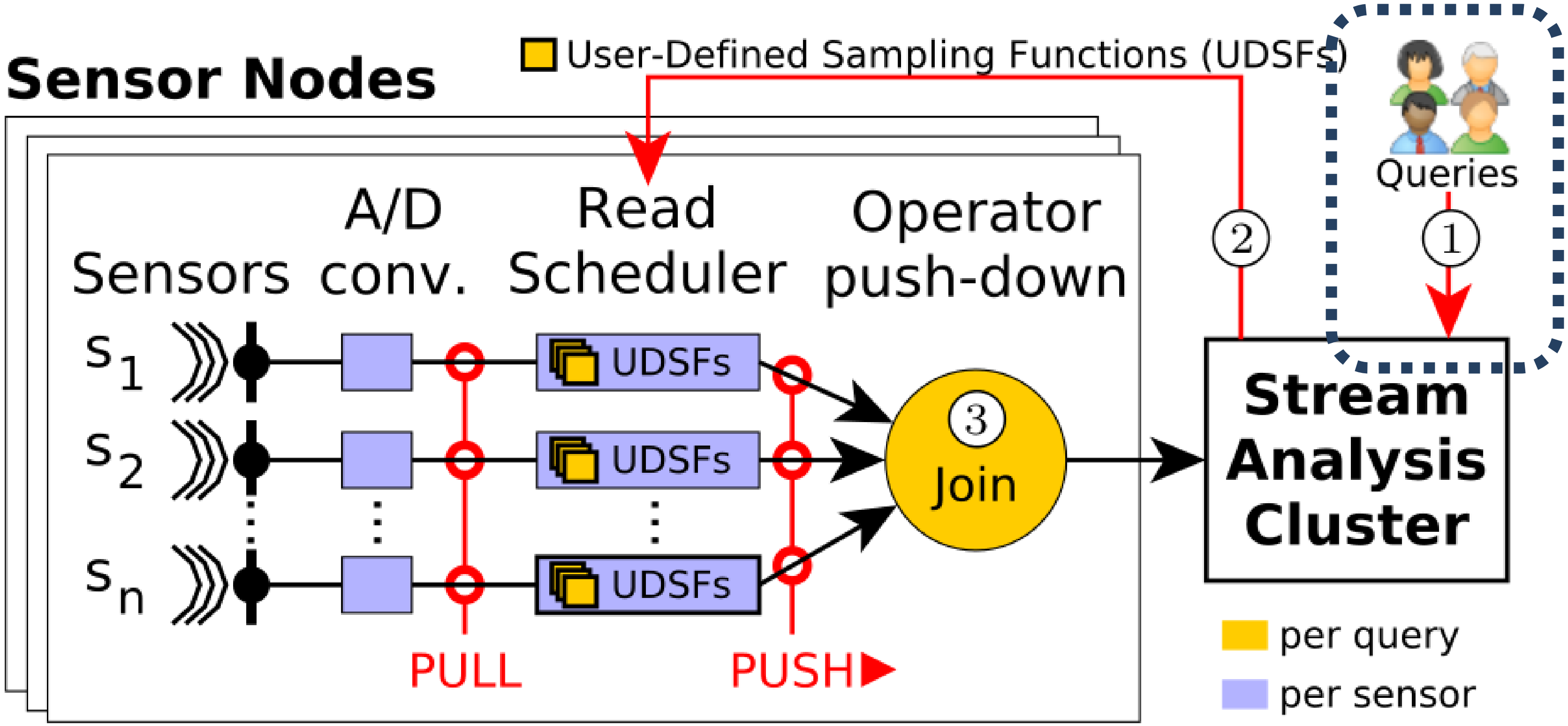
Technische Universität Berlin

German Research Center for Artificial Intelligence

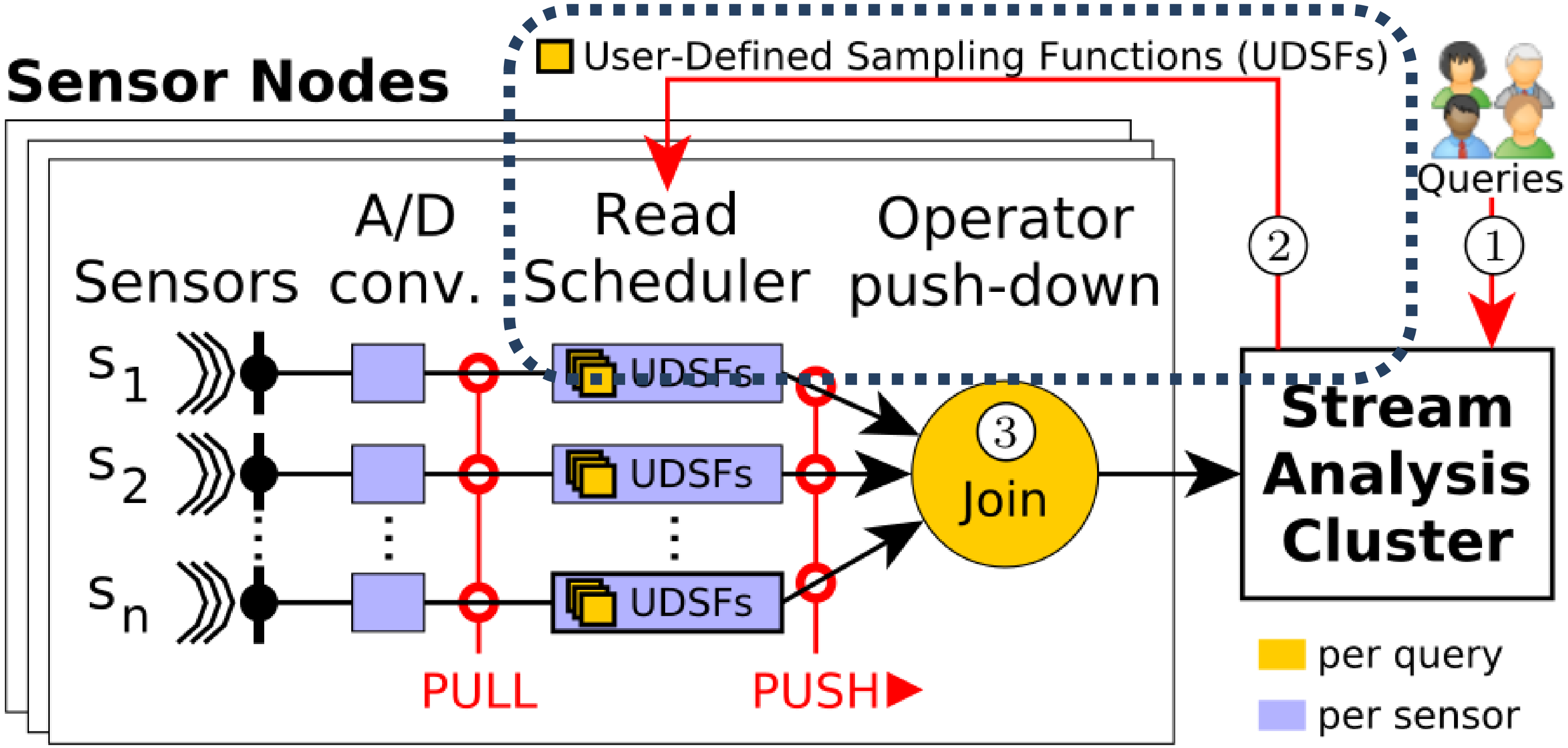
Architecture Overview



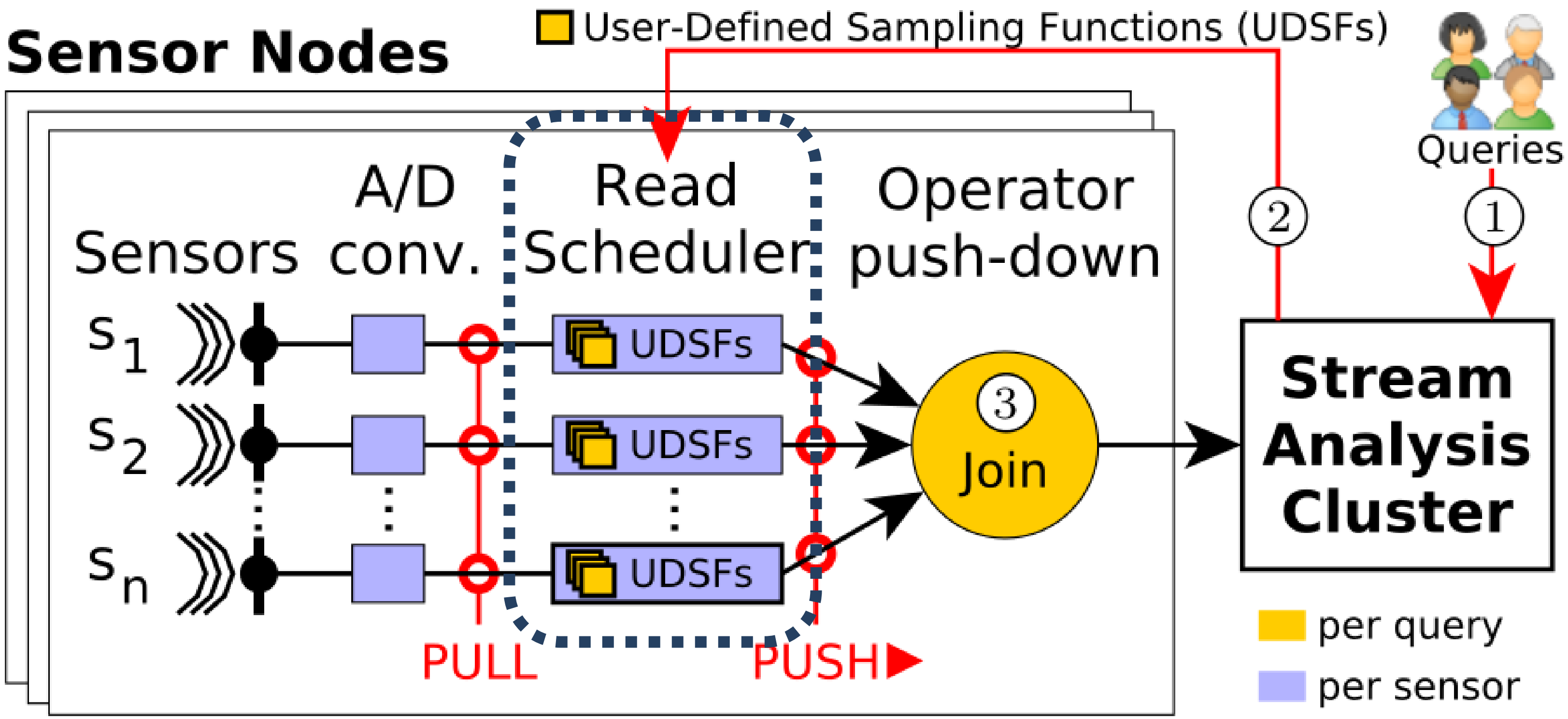
Architecture Overview



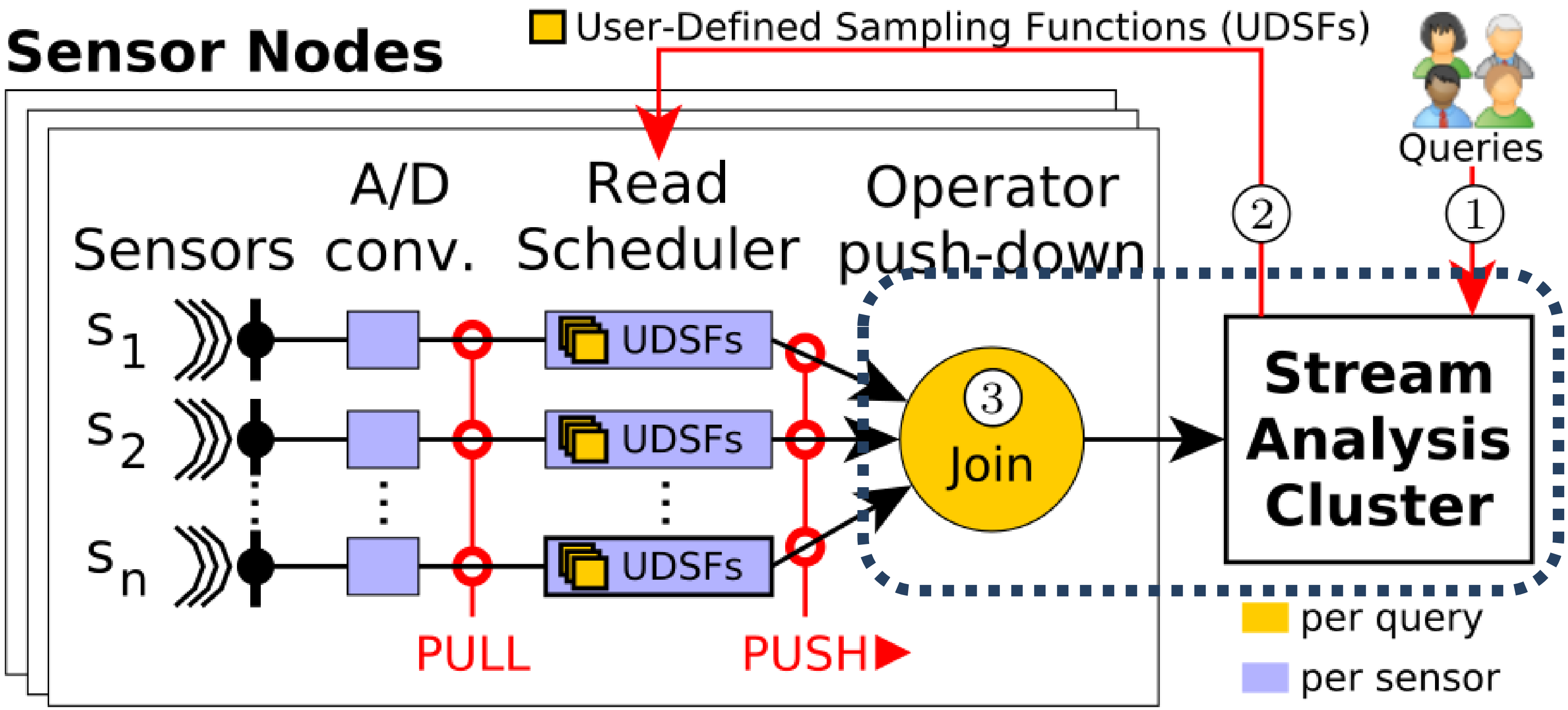
Architecture Overview



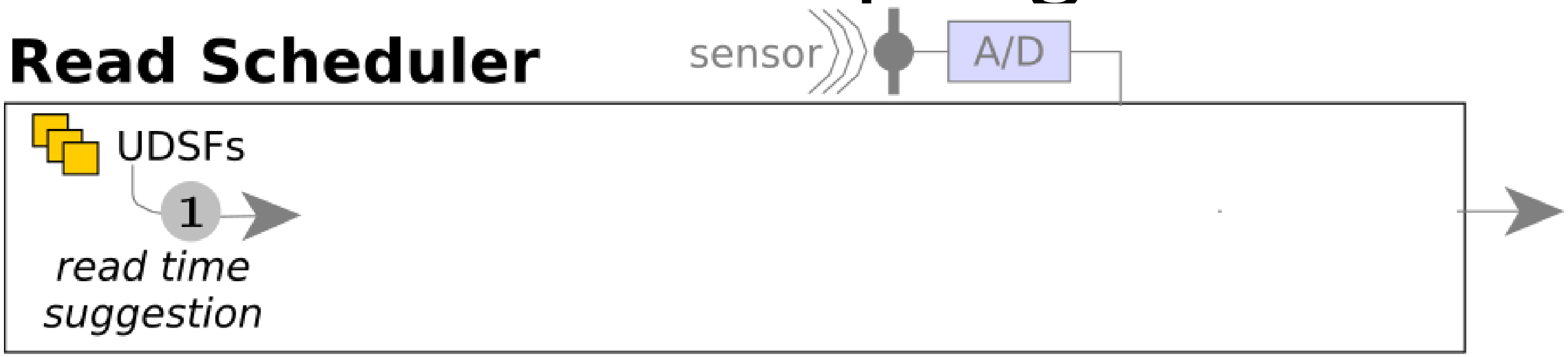
Architecture Overview



Architecture Overview

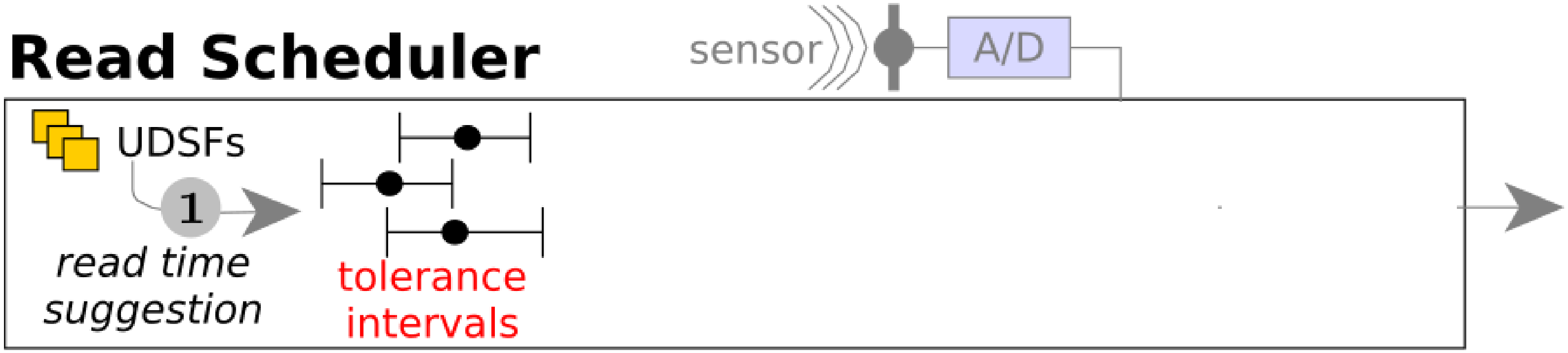


User-Defined Sampling Functions



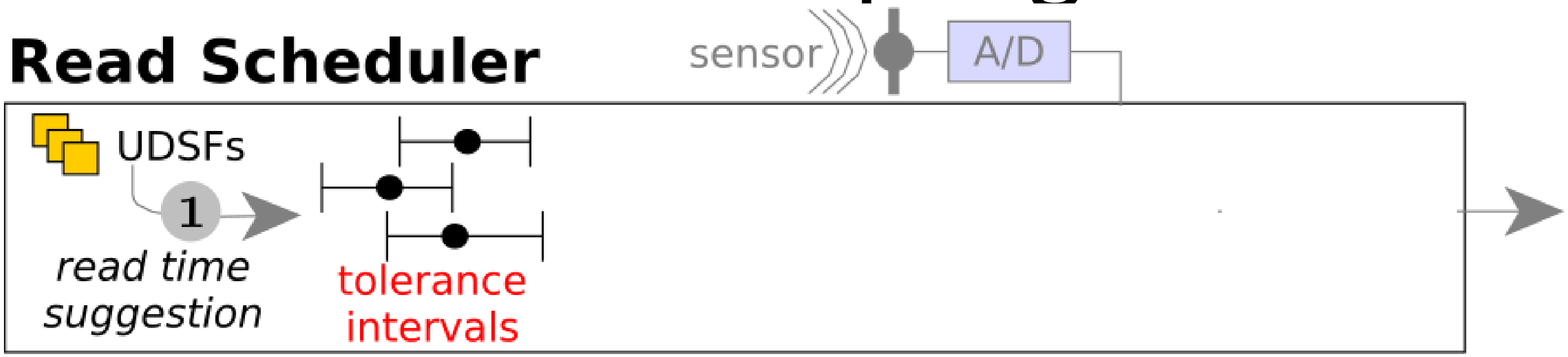
- Provide an abstraction to define the data demand of applications.
- Upon a sensor read, request the next sensor read.

User-Defined Sampling Functions



- Provide an abstraction to define the data demand of applications.
- Upon a sensor read, request the next sensor read.
- Make read time tolerances explicit.

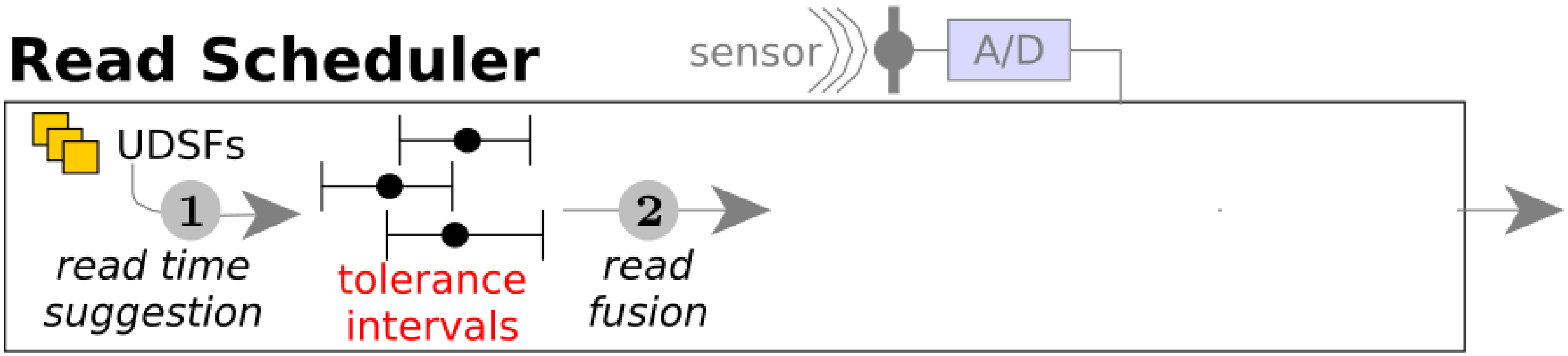
User-Defined Sampling Functions



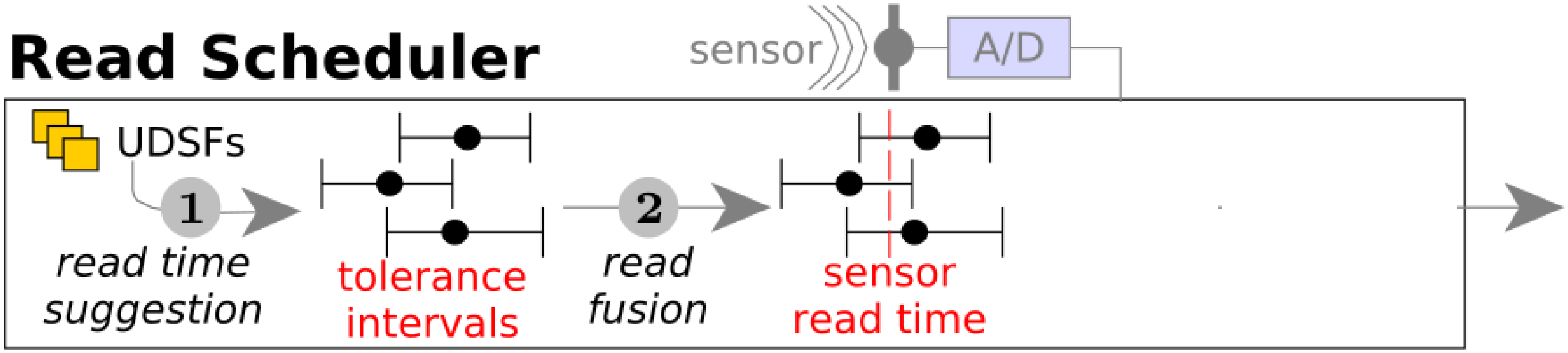
Enable adaptive sampling techniques to reduce data transmission

e.g., Adam [Trihinas '15], FAST [Fan '14], L-SIP [Gaura '13]

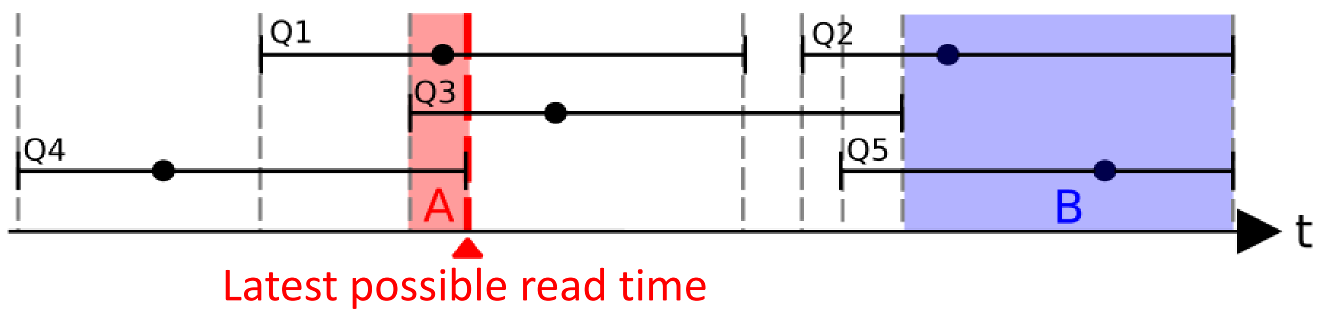
Sensor Read Fusion



Sensor Read Fusion

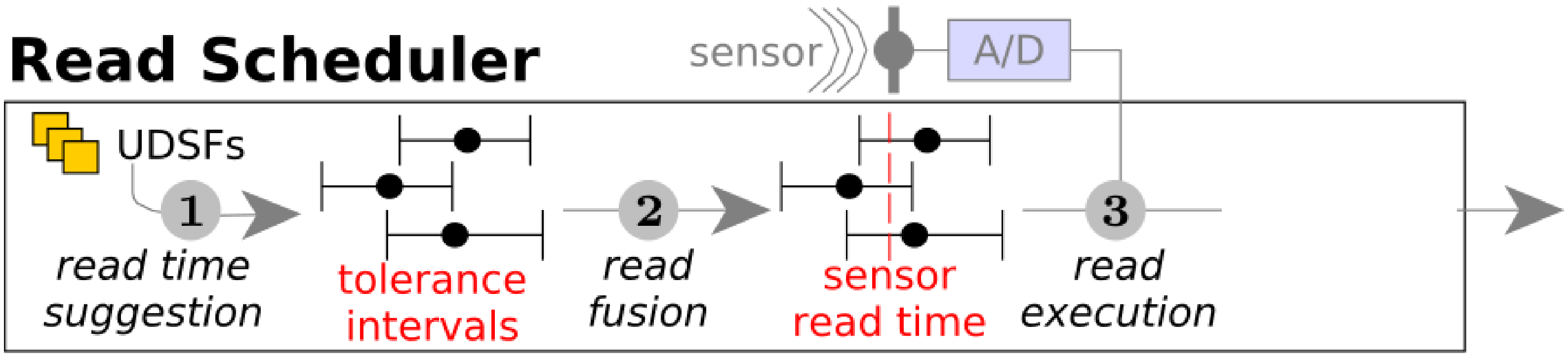


1) Minimize Sensor Reads and Data Transfer:

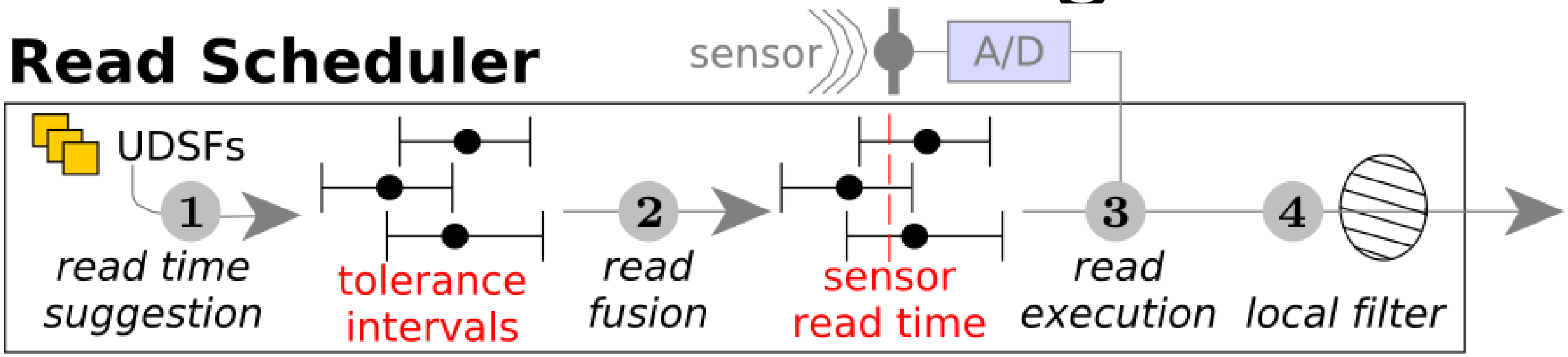


2) Optimize Sensor Read Times:

- Check the paper for all details on the read time optimizer!


Read Scheduler
sensor
A/D
UDSFs
1
read time suggestion
tolerance intervals
2
read fusion
sensor read time
3
read execution

Local Filtering



Optimized On-Demand Data Streaming from Sensor Nodes

Jonas Traub, Sebastian Breß, Asterios Katsifodimos, Tilmann Rabl, Volker Markl

Wrap-Up:

Tailor Data Streams to the Demand of Applications

- Define data demand: User-Defined Sampling Functions
- Schedule sensor reads and data transfer on-demand
- Optimize read times globally - for all users and queries

Technische Universität Berlin

German Research Center for Artificial Intelligence

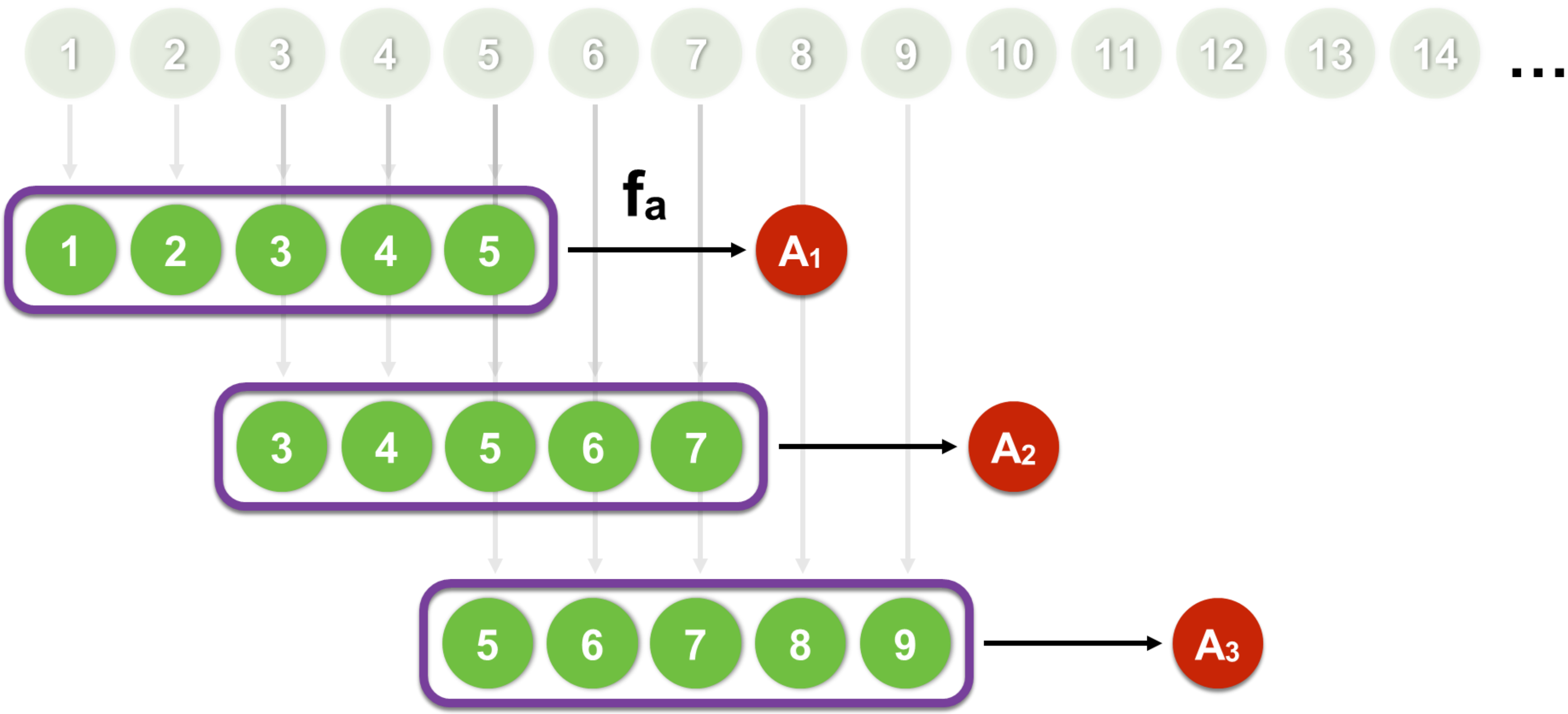
Cutty: Aggregate Sharing for User-Defined Windows

Paris Cabone, **Jonas Traub**, Asterios Katsifodimos, Seif Haridi, Volker Markl

Technische Universität Berlin

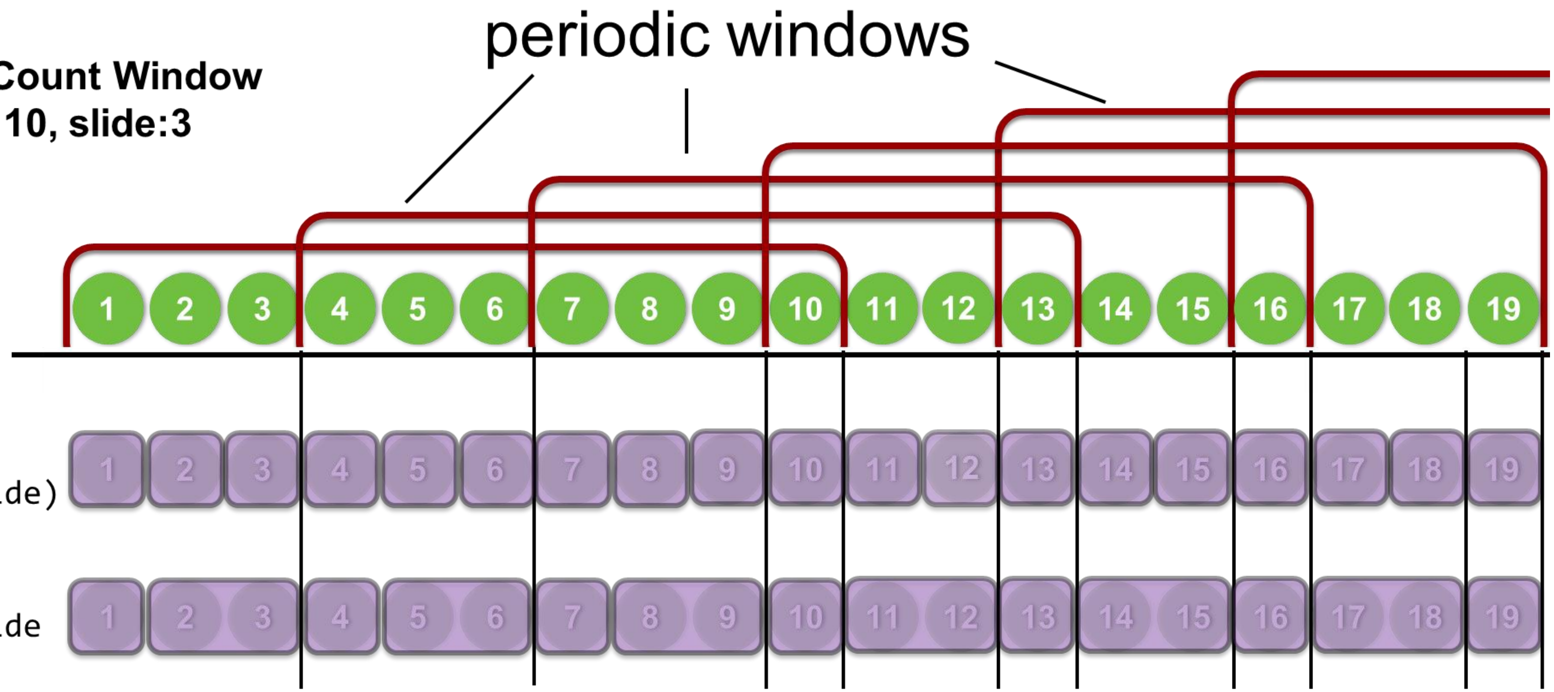
German Research Center for Artificial Intelligence

Streaming Window Aggregation



Stream Slicing

Example - Count Window
range: 10, slide:3



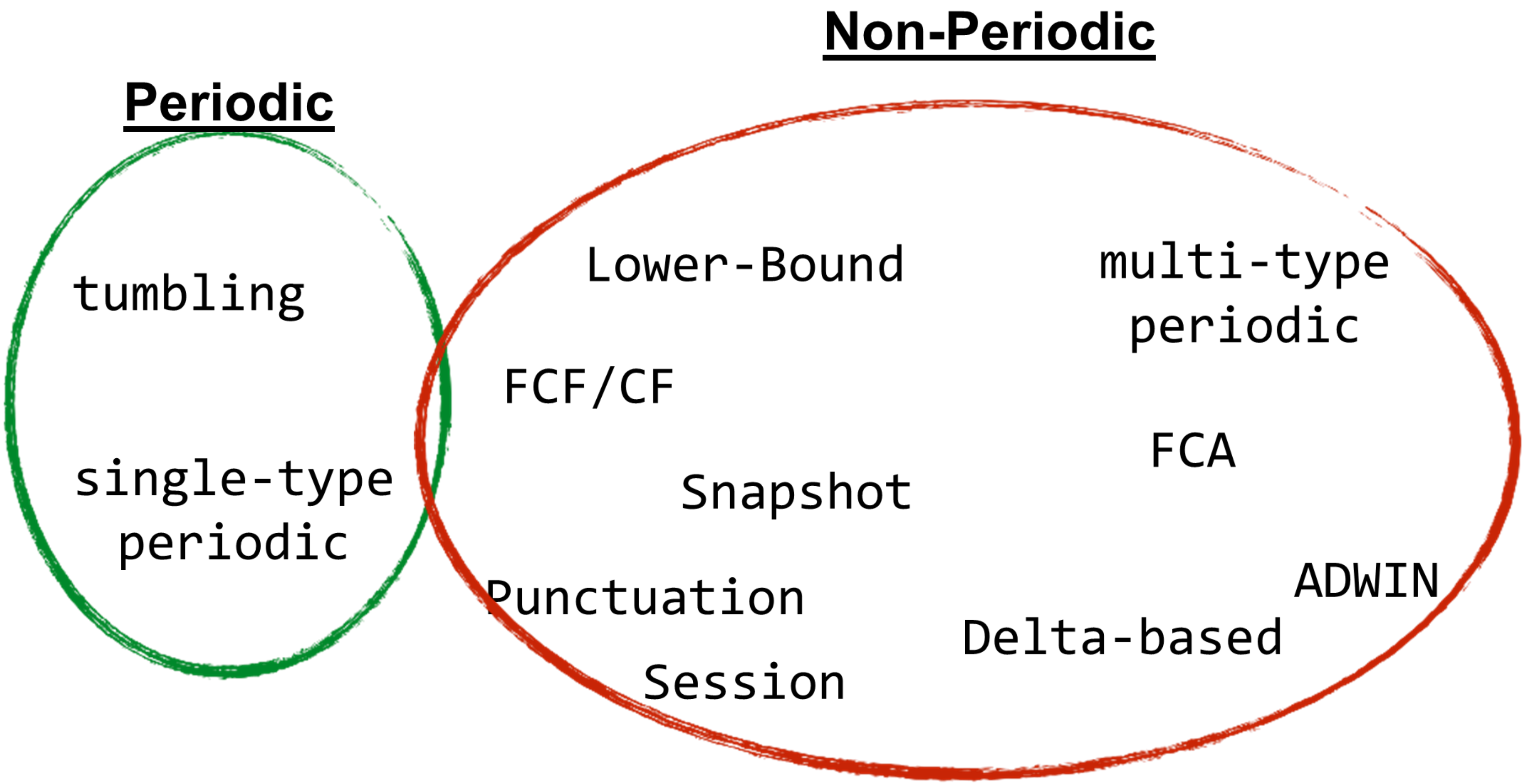
Panes

`gcd(range,slide)`

Pairs

`p2: range%slide`
`p1: slide-p2`

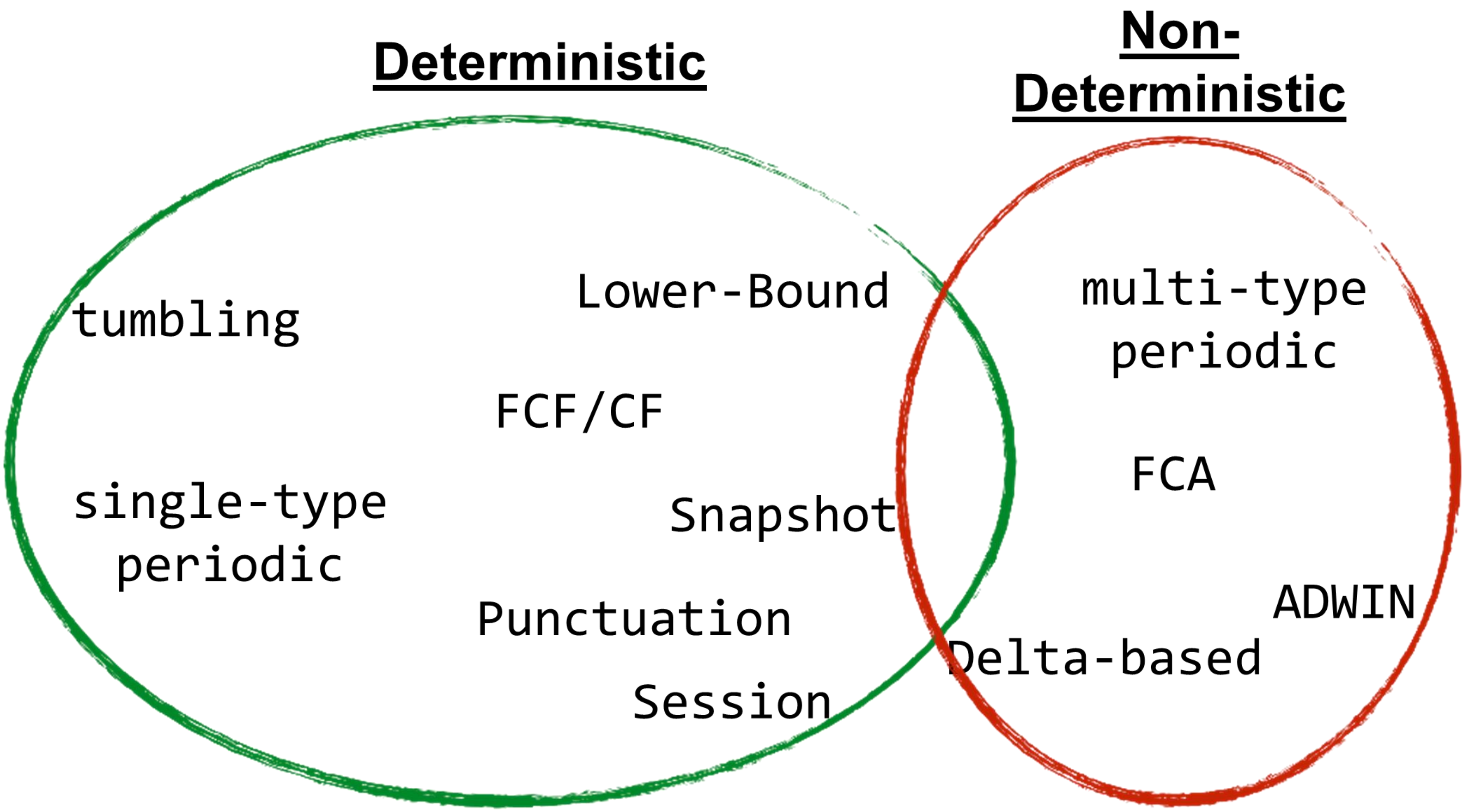
Applicability of Stream Slicing



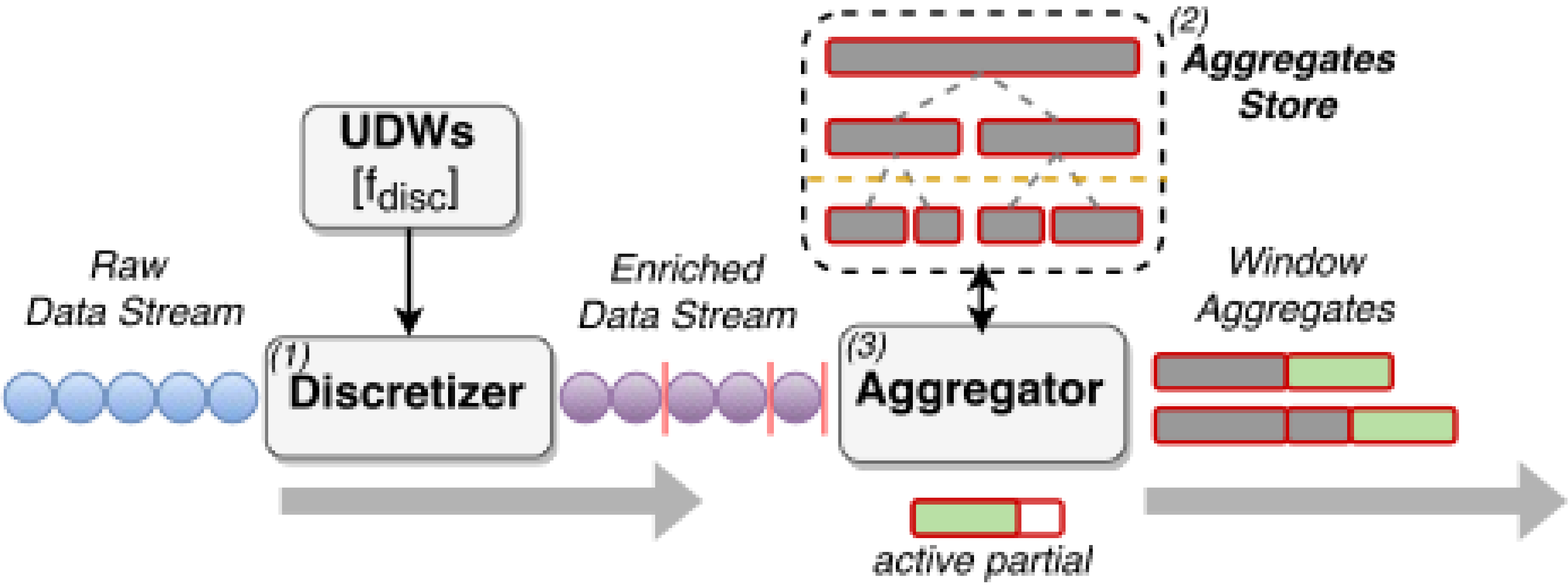
No slicing for the majority of window types

Can we do better?

Yes, we can do better!



Cutty Overview



Cutty: Aggregate Sharing for User-Defined Windows

Paris Cabone, **Jonas Traub**, Asterios Katsifodimos, Seif Haridi, Volker Markl

Wrap-Up:

Enable Stream Slicing beyond Simple Tumbling and Sliding Windows

- Cutty enables Stream Slicing for a broad class of windows
- Cutty combines Stream Slicing, On-the-fly Aggregation, Aggregate Sharing, and Aggregate Trees

Technische Universität Berlin

German Research Center for Artificial Intelligence

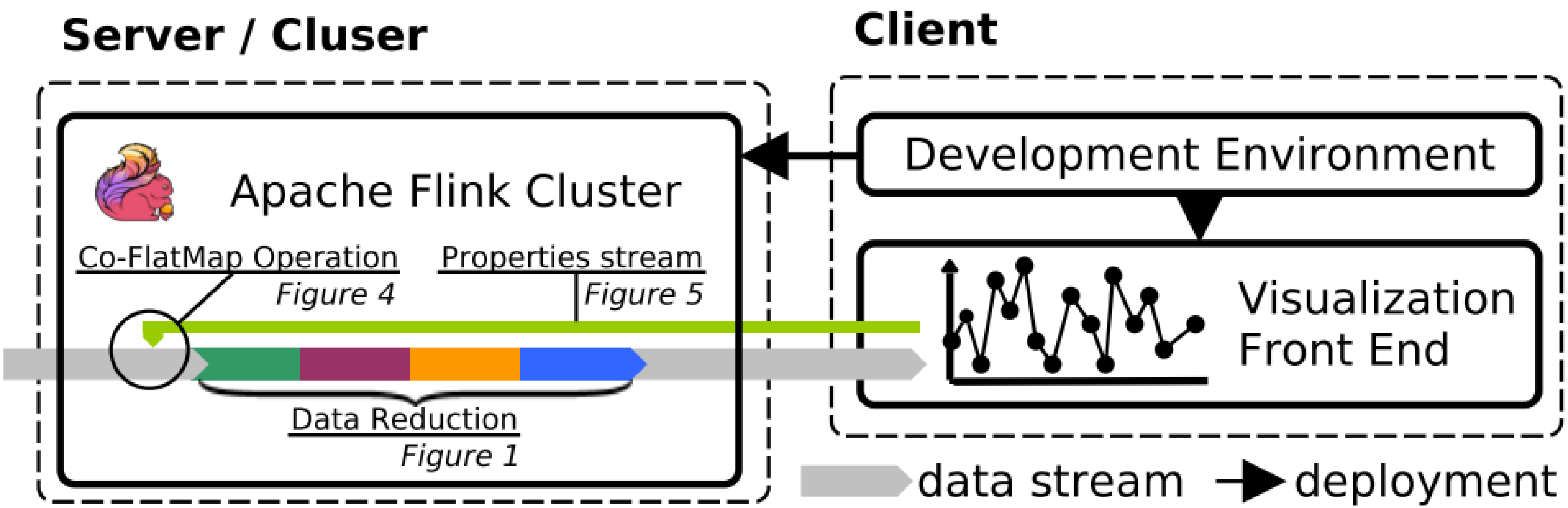
I^2: Interactive Real-Time Visualization for Streaming Data

Jonas Traub, Nikolaas Steenbergen, Philipp Grulich, Tilmann Rabl, Volker Markl

Technische Universität Berlin

German Research Center for Artificial Intelligence

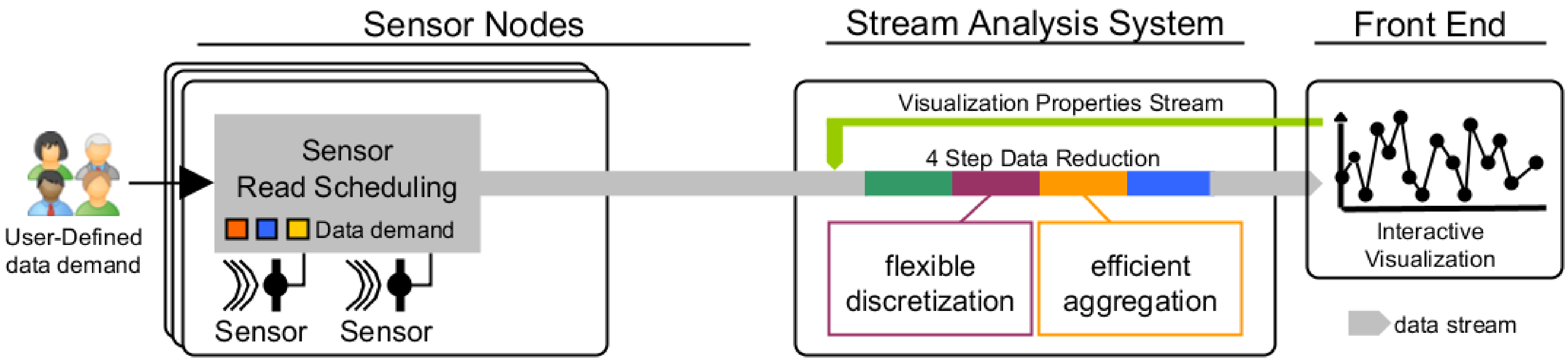
Architecture Overview



Check out our Flink Forward Talk

youtube.com/watch?v=JNbq239JkK4

The Big Picture



Optimized On-Demand Data Streaming from Sensor Nodes
Traub et al.; ACM SoCC'17

I²: Interactive Real-Time Visualization for Streaming Data
Traub et al.; EDBT'17

Cutty: Aggregate Sharing for User-Defined Windows
Carbone et al.; CIKM'16