HSTREAM — Big Stream Processing on IOT Environments

Genoveva Vargas-Solar
CNRS-LIG-LAFMIA

Javier A. Espinosa-Oviedo
BSC-LAFMIA

Enrique Arriaga-Varela**
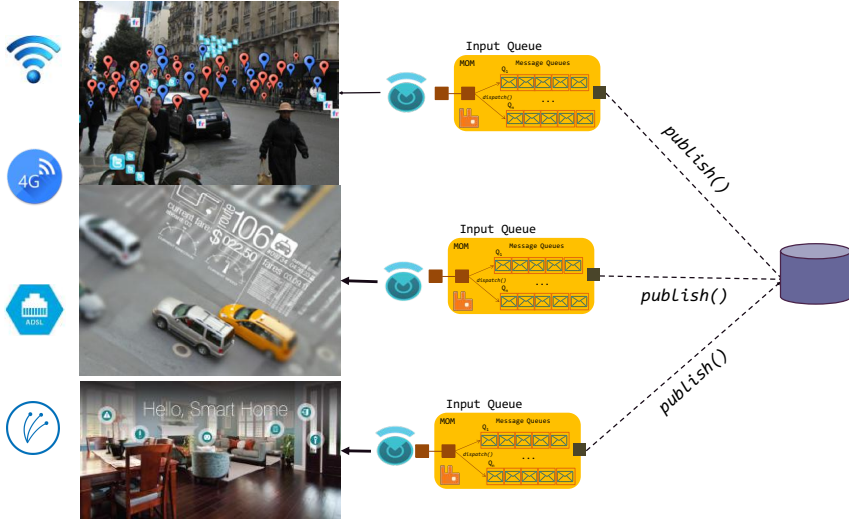UAG

Christine Collet
Grenoble INP

Analysing connectivity for connected Society

Data Acquisition
Continuous monitoring of connected society

Data Storage
Archival of data as files

publish()

Attribute	Description
Time	Number of seconds elapsed after EPOCH
Asnum	Internet provider' ID.
Client_country	Country where the test was conducted.
Connect_time	Round-Trip Time
Download_speed	Download speed (bytes/secs).
Lat	Latitude of the GPS location
Latency	Delay between the sent and reception
Lon	Longitude of the GPS location
Mlabservername	Name of the server
Platform	User operative system
Upload_speed	Upload speed (bytes/secs).

Neubot — Network measurements from the edges
www.neubot.org

Project **S2EUNET** (EU-FP7 IRESES)
LIG, LAFMIA, Politecnico di Torino
http://s2eunet.org

o **Connection speed of one internet provider**
Am I receiving the network speed I am paying for all the time?
o **Internet availability with different Internet providers**:
Which are the periods of the day in which I can upload/download files at the highest speed using different network providers?

Big Stream Processing System

Stream Operator *(core concept)*
Software component in charge of processing data

Operator Architecture
Queue, Fetch, Buffer, Sink

```
EVERY       2 minutes give me the MAX value of download_speed
STARTING    8 minutes ago
FROM        Influx data Neubot
            series DownloadSpeed
            queue Neubot
```
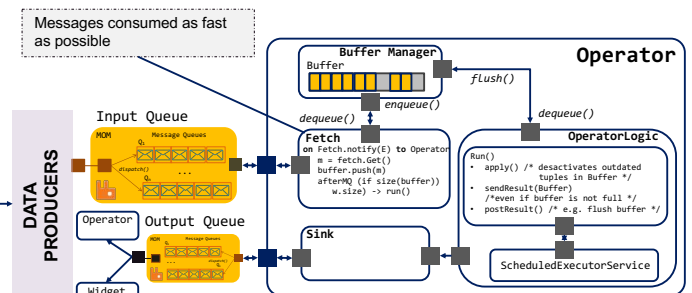
```
5 ROWS:
Timestamp                 Count      Value
2017-10-03 20:52:00.0     7          108437.0
2017-10-03 20:54:00.0     12         65178.0
2017-10-03 20:56:00.0     12         113911.0  << result
2017-10-03 20:58:00.0     12         107199.0
2017-10-03 21:00:00.0     4          17327.0

INITIAL max VALUE = 113911.0
```

Messages consumed as fast as possible


```
def queryToHistoric(function: String, value: String,
                    startTimeInMillis: Long, endTimeInMillis: Long,
                    groupByTimeNumber: Int, groupByTimeTimeUnit: String)
    : List[List[Object]]
```

Big Data operators

SPEPSIoT

Filter	
Input	
Join	
Merge	
Sort	
Data sum	
Data average	

BOP

InfluxDB

SpepsIoT Component
+getComponentID()
+getName()
+getIcon()
+start()
+stop()
+bind()
+unbind()
+...()

DB Interface

SPARK — APACHE

historical tuples

Operator Interface

tuples (stream)

Operator

Driver Program — SparkContext — Cluster Manager

Worker Node — Executor — Cache — Task — Task

LIG — Laboratoire d'Informatique de Grenoble