A Brief History of Time with Apache Flink

Real time monitoring and analysis with Flink, Kafka and HBase

Thomas LAMIRAULT

Mohamed Amine ABDESSEMED

Flink Forward 2016

Berlin, September, 12th, 2016

bouygues TELECOM

The speakers

Thomas LAMIRAULT

- Bigdata Software engineer @ Bouygues Telecom since 2015
- A Flink master enthusiast

@thomaslamirault

Mohamed Amine ABDESSEMED

- Software engineer & solution architect @ Bouygues Telecom since 2013
- A Flinker since the early beginnings

@AminouvicTweets

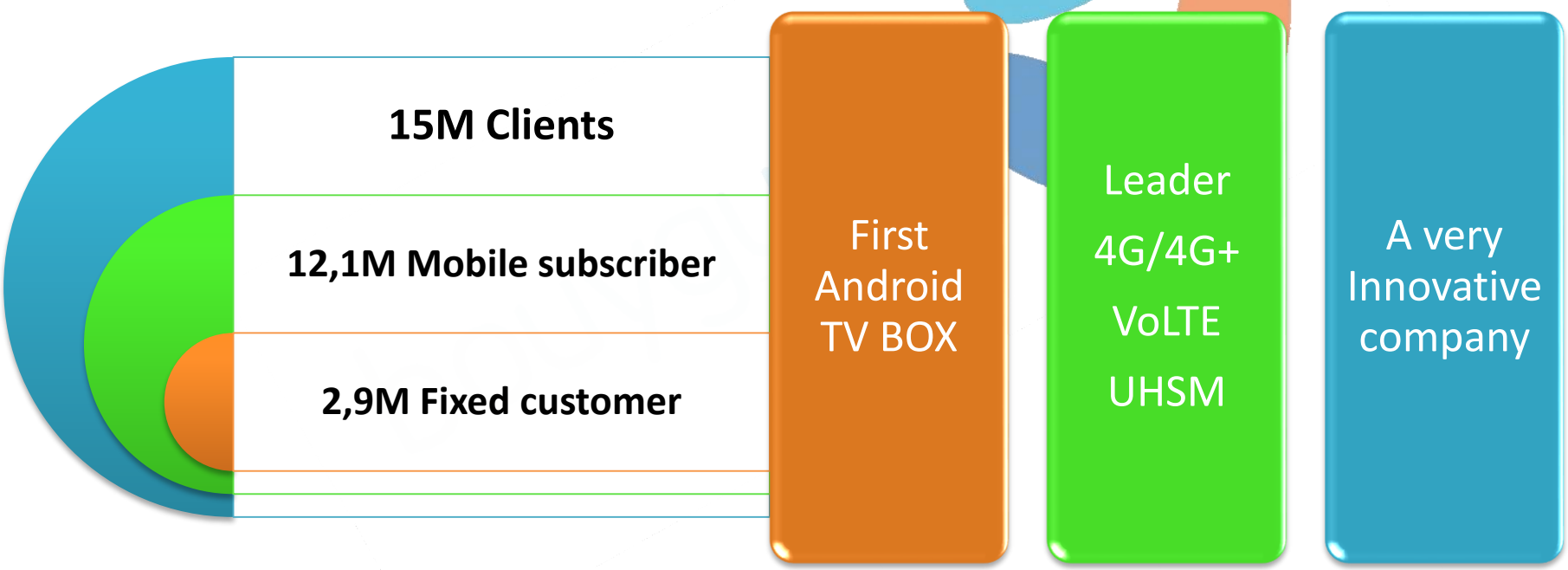
bouygues TELECOM

Outline

- Who is Bouygues Telecom
- Data Value and Streaming Analytics
- Use case
- Challenges
- Streaming analytics with Flink
- Results

WHO IS BOUYGUES TELECOM ?

Mobile . Fixed . TV . Internet . Cloud



we love technology

bouygues TELECOM

WHO IS BOUYGUES TELECOM ?

4G **4G+** ***Ultra High Speed Mobile***

A nationwide network covering 75% of the population

Commercial launch in June 2014
in France's major towns and cities, with speeds up to two times faster than 4G.

The year of Ultra High Speed Mobile
Since 2015, Ultra High Speed Mobile has provided speeds of close to 300 Mbit/s thanks to the aggregation of three frequency bands in some major cities, including Lyon and Chartres. In 2016, this roll-out will continue in the major towns and cities of France.

Coverage

2,000 network experts

15,000 mobile sites in France

and roaming agreements in more than 272 destinations, of which 55 for 4G

bouygues TELECOM

LUX: Logged User eXperience

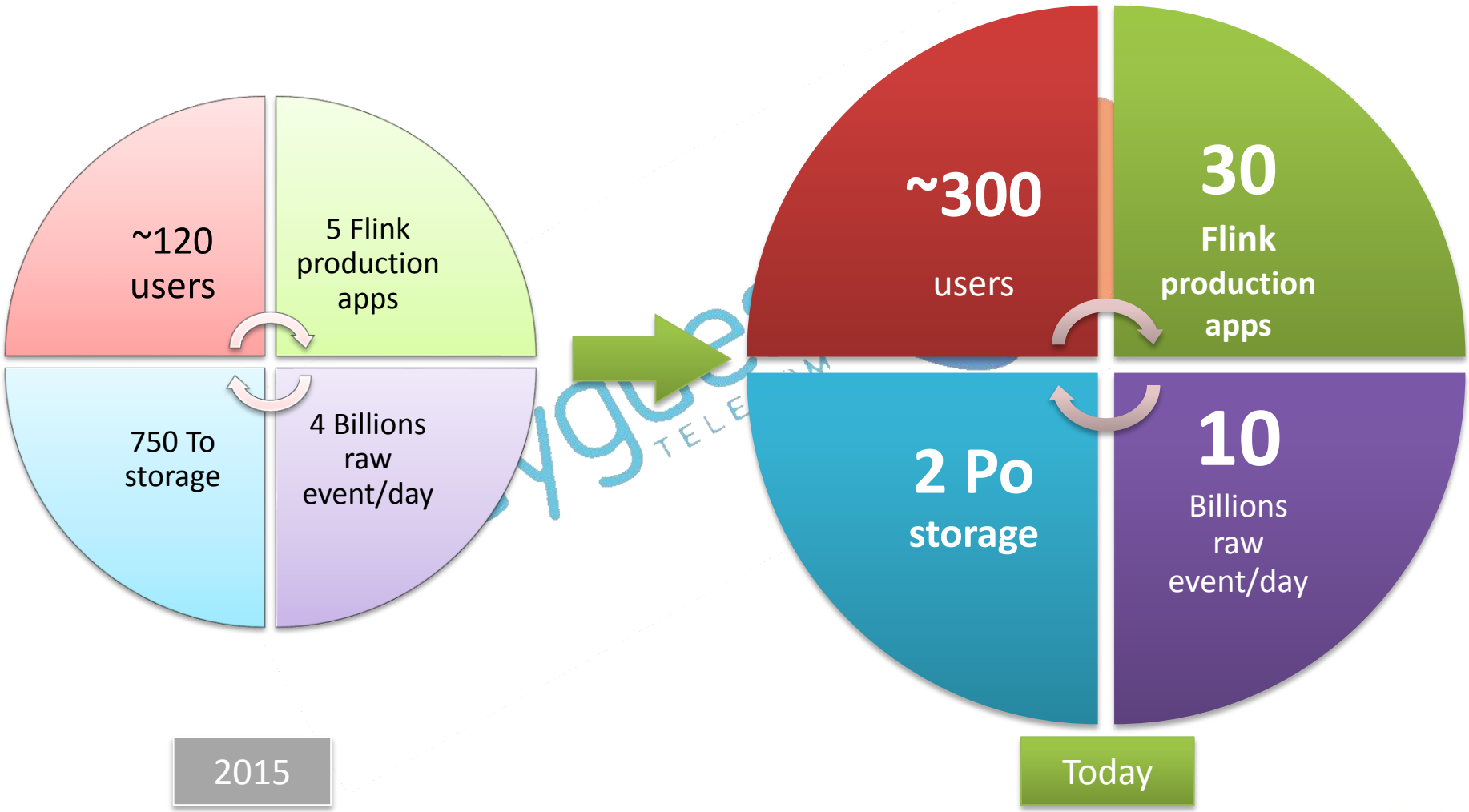
Mobile QoE

- Our Big Data platform
- Produce Mobile QoE indicators based on massive network equipment's event logs (**Billions event/day**).

- Goals:
  - QoE (User) instead of QoS (Machine)
  - Real-time Diagnostic (<60' end-to-end latency)
  - Business Intelligence
  - Reporting
  - Real-time alarming

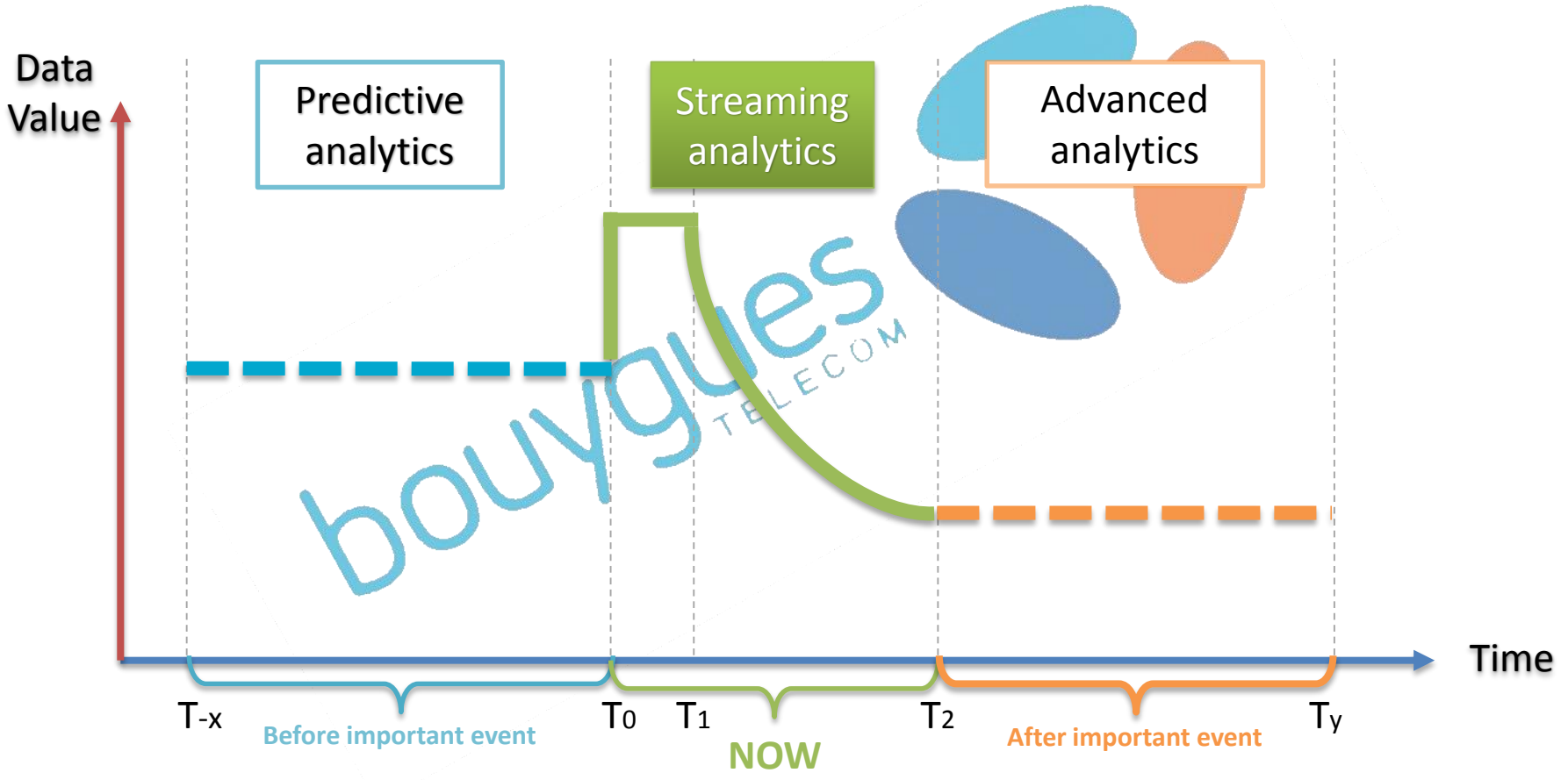
LUX

In Numbers



bouygues TELECOM

Analytic Data value



DATA IS MOST VALUABLE NOW !

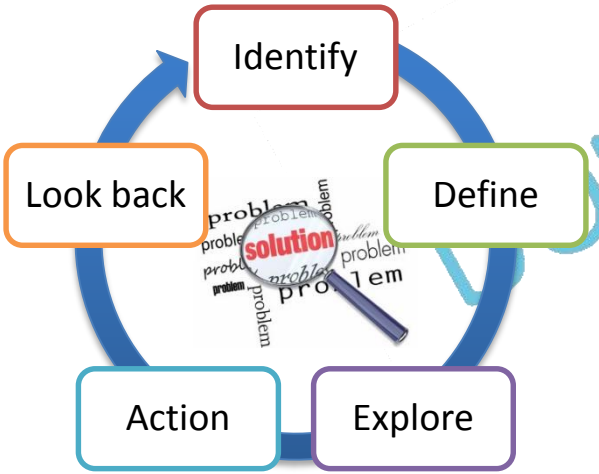
bouygues TELECOM

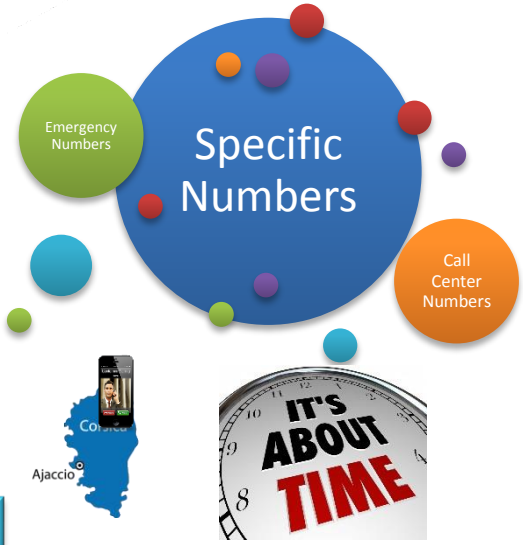
Analytic Data value

- Data is most valuable when made available as soon as important events occur.
- Get the most of Data
  - Collect data **fast**.
  - (Pre)Process it **fast**.
  - Analyze it and create added value to act **faster**!

The use case

GNOC





The BIG picture isn't always significant

Global counts have no sense !

bouygues TELECOM

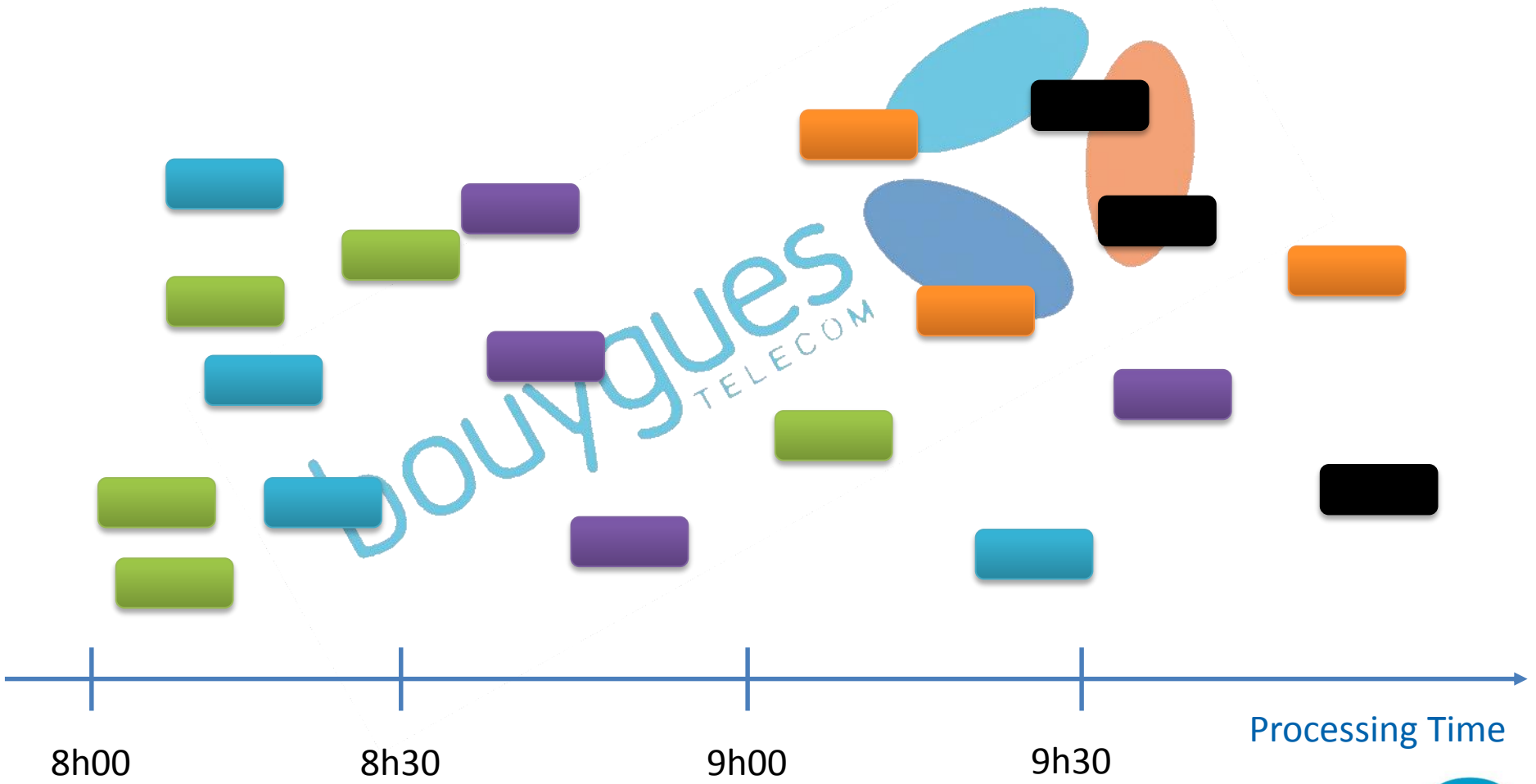
The use case

- A simple and valuable use case
- Need to analyze the entire call traffic :
  - Considering multiple aggregation axes
  - Fine grained analysis
  - Detect when something is happening somewhere in real-time
  - Compare with historical values trends

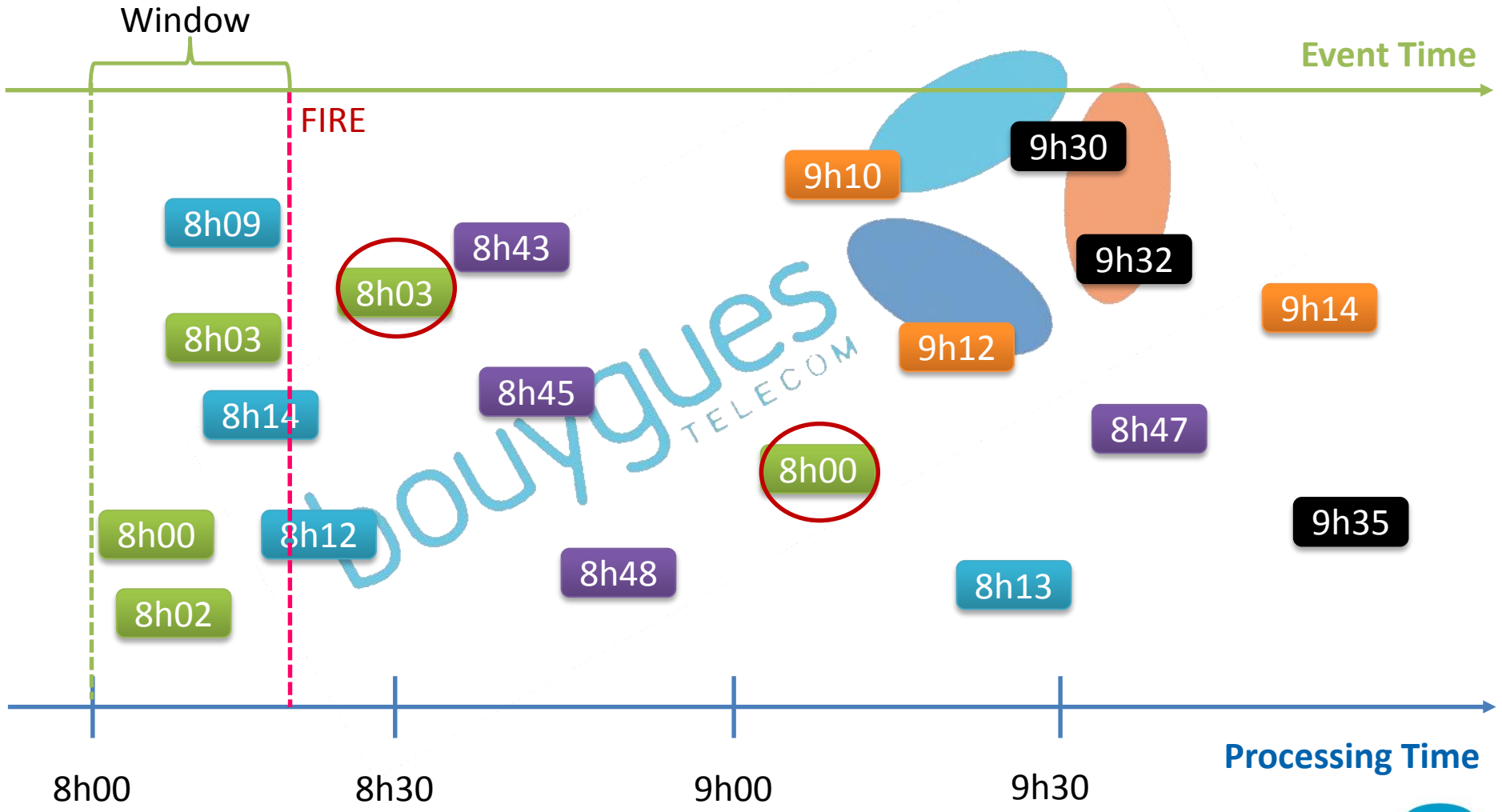
Challenges

- Low latency & streaming fashion counters
- Quickly available KPIs = value
- Massive amounts of data + peak loads
- Reliability
- Multiple flow correlation
- Time management:
  - Out of order & late events ➔ our worst enemies
  - Flexible window management
  - Specific watermark emission

Streaming Analytics Time Management



Streaming Analytics Time Management



Streaming Analytics with Flink

Built-in windowing functionalities

- Custom Watermark extractor
- Custom Triggers for lateness management
- Custom Key extractor

Stateful Streaming

- Checkpointing
  - Fault tolerance
- Savepoints
  - Update without data loss

Streaming analytics with Flink

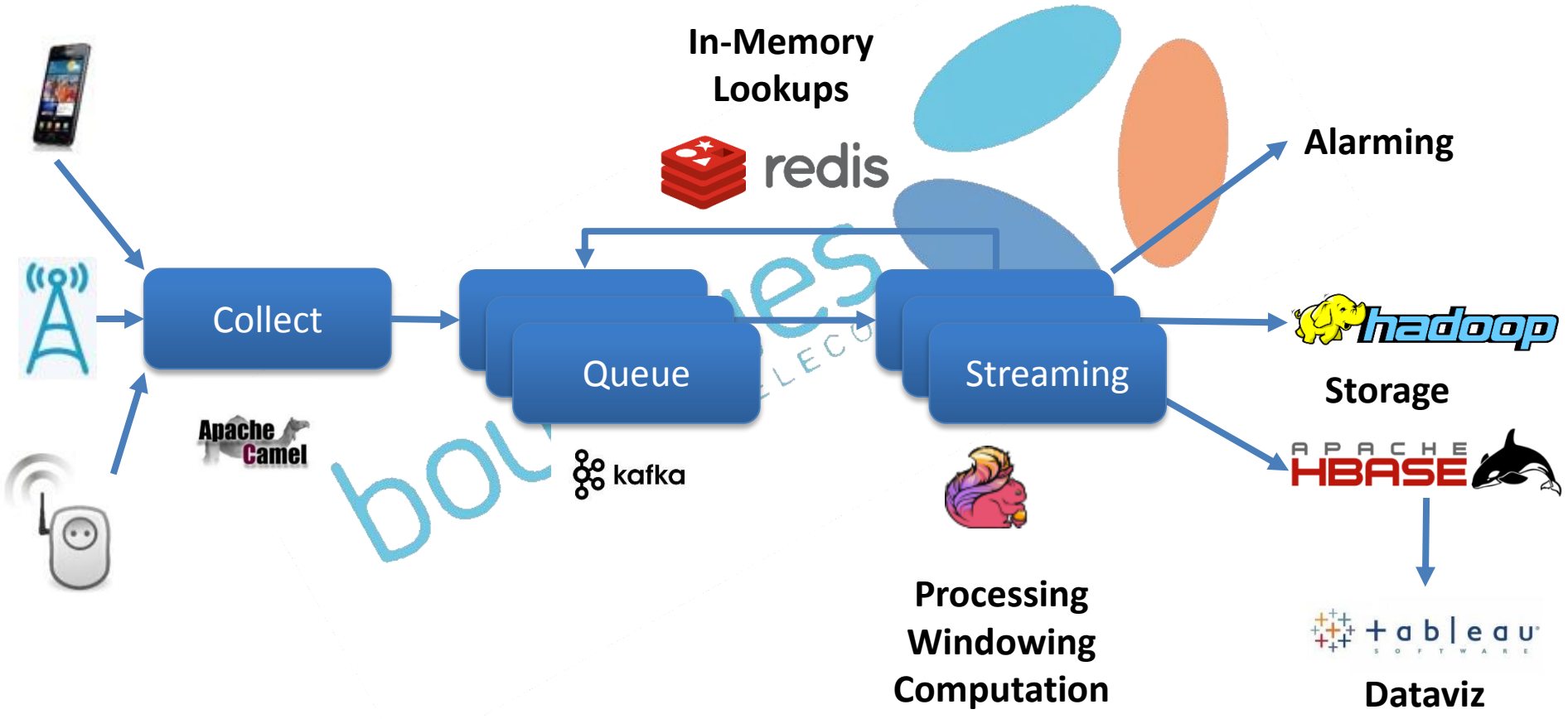
Performance

- High throughput
- Low latency
- Excellent memory management

Flexible window management

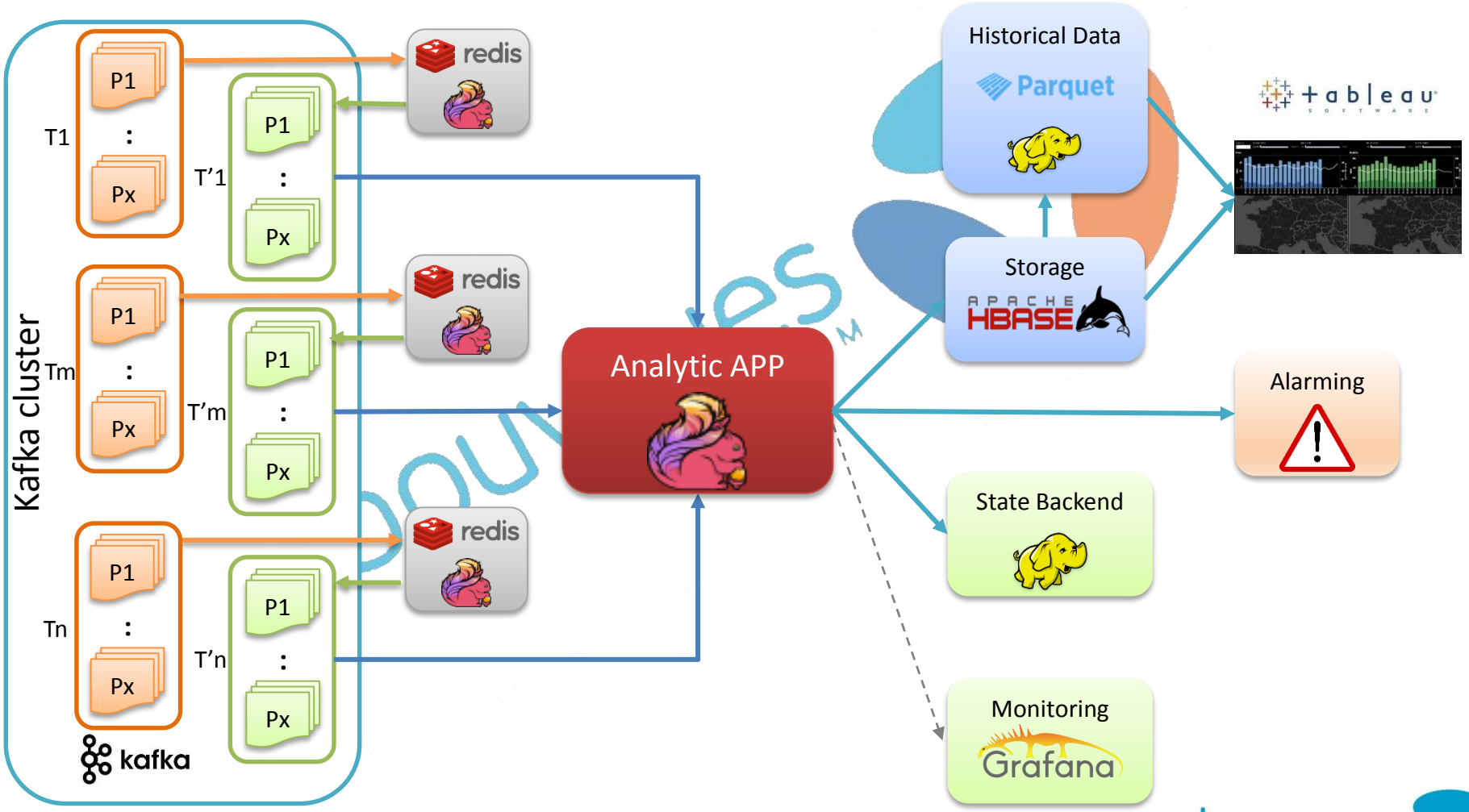
- Tumbling
- Sliding
- Session

High Level Architecture



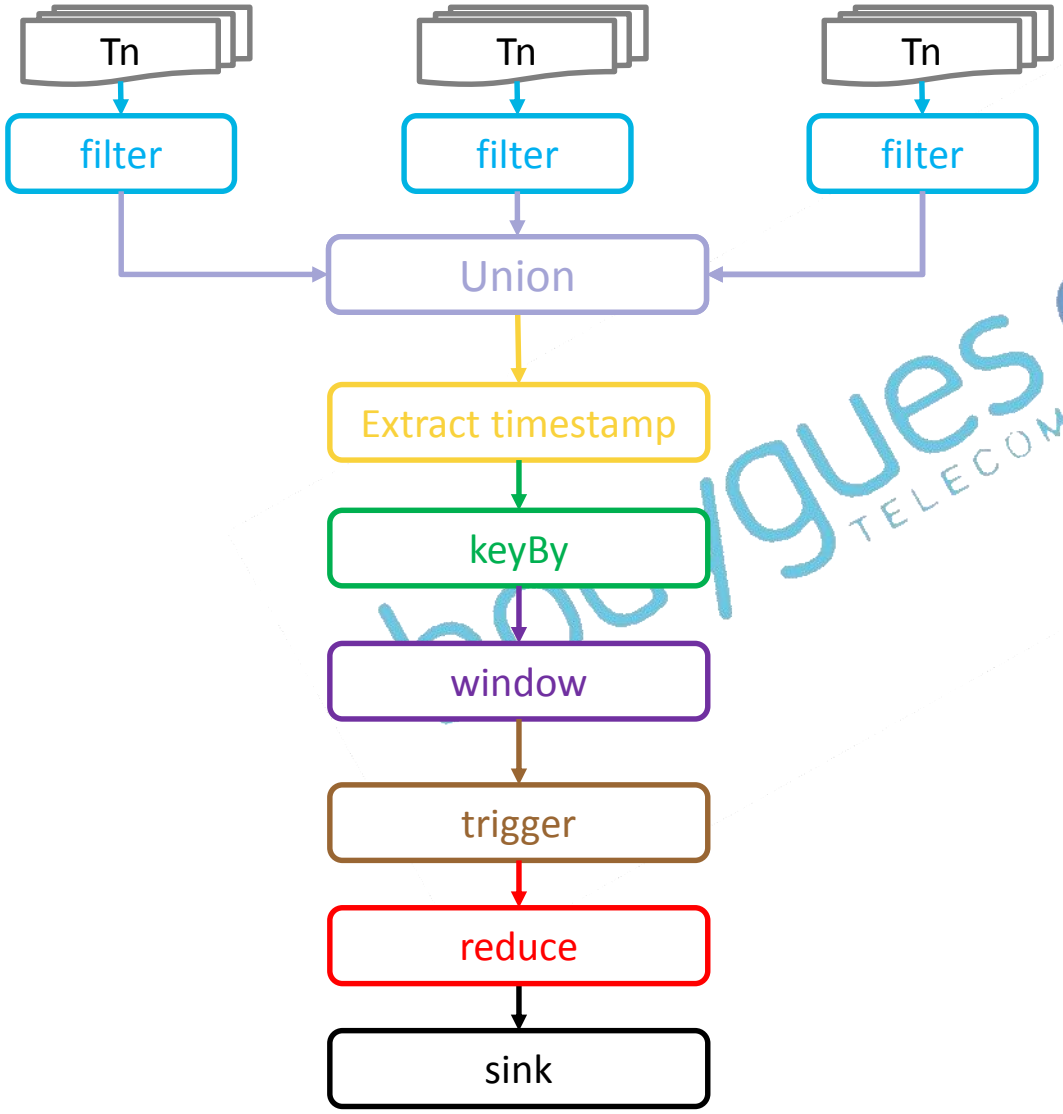
bouygues TELECOM

Architectural details



bouygues TELECOM

Streaming application Details



Multiple Input Topics

Keep only records that interest us

Correlate input flows

Extract relevant event timestamp, with custom watermark emission

Group by considered aggregation axes

Tumbling windows

Custom trigger, allow configurable late data and manage lagged data

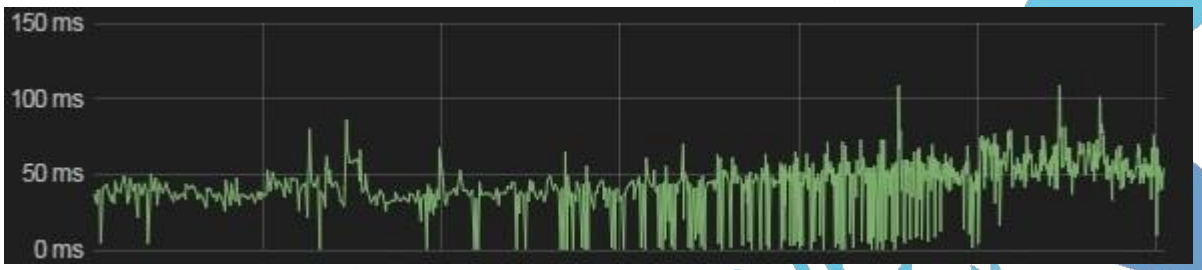
Produce KPIs

Write Data into HBase

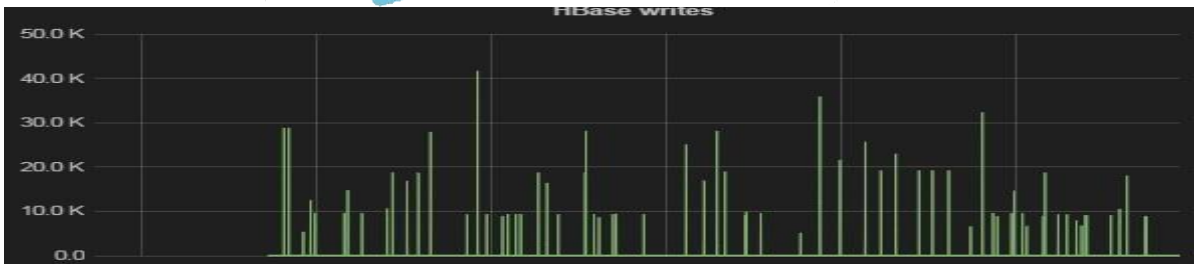
The results

Production metrics

- Low latency (<100ms)

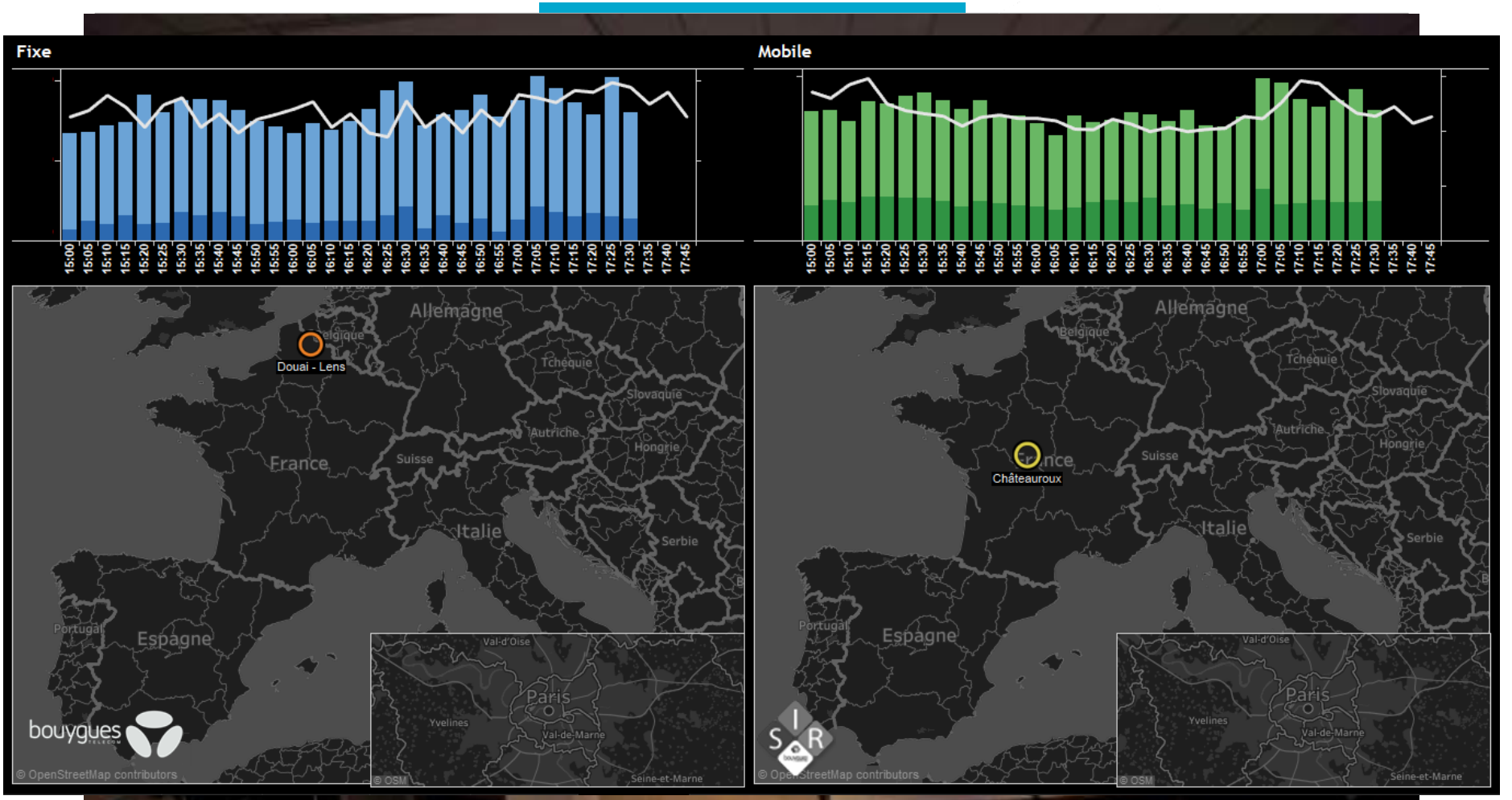


- Input : Up to ~80.000 events/sec
- Output : Produce ~40.000 KPI/window



bouygues TELECOM

The results
Dataviz



Benefits

- Monitor and improve customer experience
- Reduced incident detection time
- Help GNOC alarm prioritization based on customer experience
- Reduced operating costs

Difficulties

- Massive amounts of data in both **input** and **output**
- Savepoint/Checkpoint cost
- HBase analytic limitations
  - Read vs Write
  - Long Scan
- Massive out of order events

What's Next

- Flink + Kudu
- Async/Incremental checkpoints
- Flink CEP
- Streaming SQL
- Flink applications monitoring & industrialization.

Questions ?

bouygues TELECOM