

# Managing Big but also Fast data with CEP

Gibert Philippe – Orange/OLNC  
June 26<sup>th</sup> 2013 , Forum Teratec – Big Data & HPC





## Agenda

- **Background**
- Why Big Data?
- Why Fast Data?
- Why Big + Fast ?
- Key issues
- Use Cases
- Work in progress
- Conclusion
- References



## Background

- the Orange Group (March 2013):
  - 170,000 employees worldwide (104,000 employees in France)
  - 172 million mobile customers
  - 15 million broadband internet (ADSL, fibre) customers
- entity OLNC/OLPS → R&D:
  - Products & Services Development (LBpro , PABX, M2M )
  - Research
    - MW & Platforms for Big & Fast Data,
    - Complex Event Processing (CEP)
    - Event oriented platforms



## Telecom Context

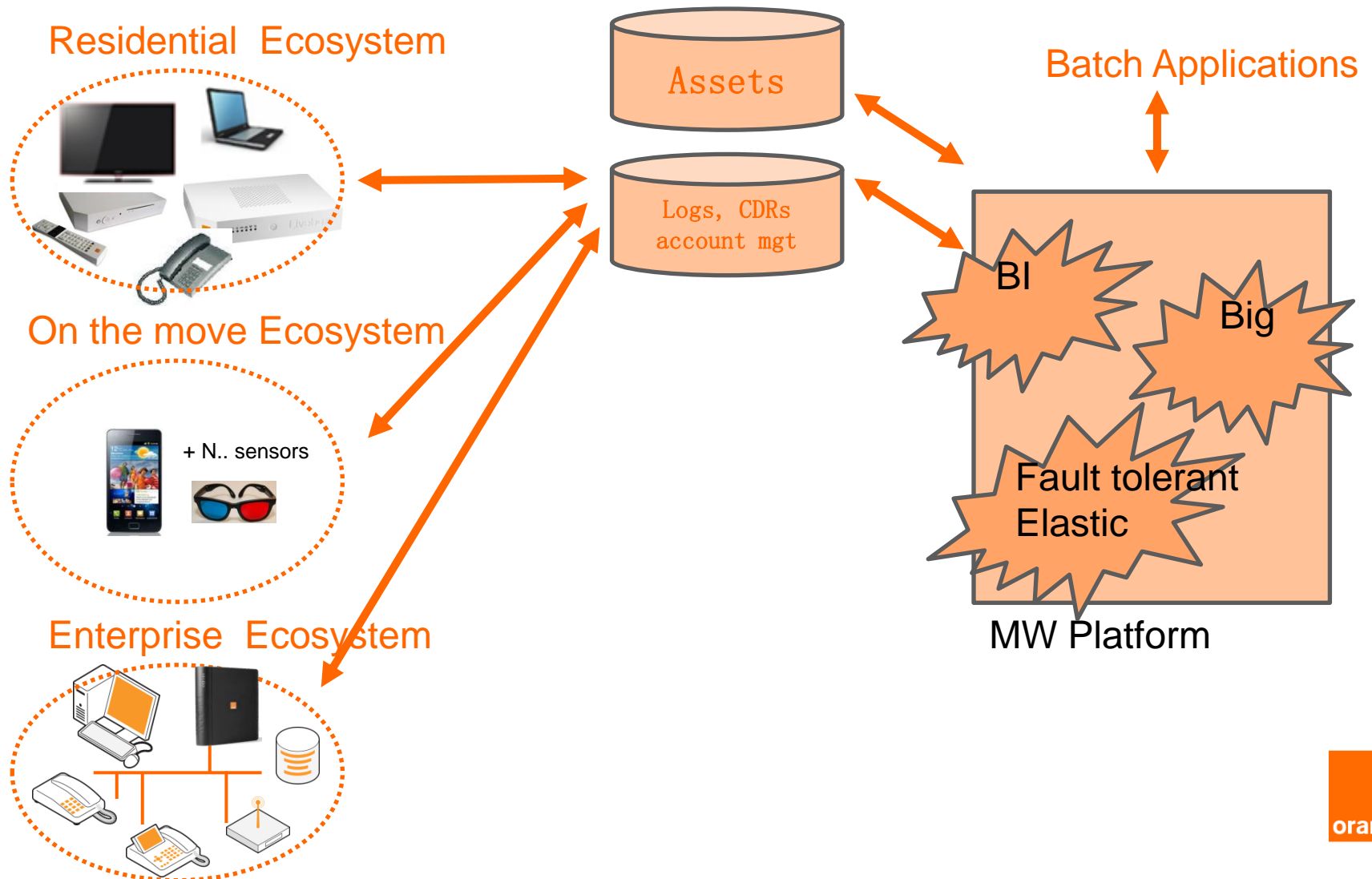
- tons of Data and Events :
  - generated by Telco's Products and Services
    - billing, CDRs, internet, OTTs, Social Networks
    - smartphones sensors, M2M devices, RT devices
- quick evolution in a few years :
  - traditional SGBD → Big Data MW (Open Source, large community)
  - traditional Analytics → Real Time Analytics
- challenge
  - adapt the legacy Platforms, Services and Applications to these Big Data and RT constraints
  - deliver the famous (infamous ?) Added Value ...



## Agenda

- Background
- **Why Big Data?**
- Why Fast Data?
- Why Big + Fast?
- Key issues
- Use Cases
- Work in progress
- Conclusion
- References

# Why Big data?

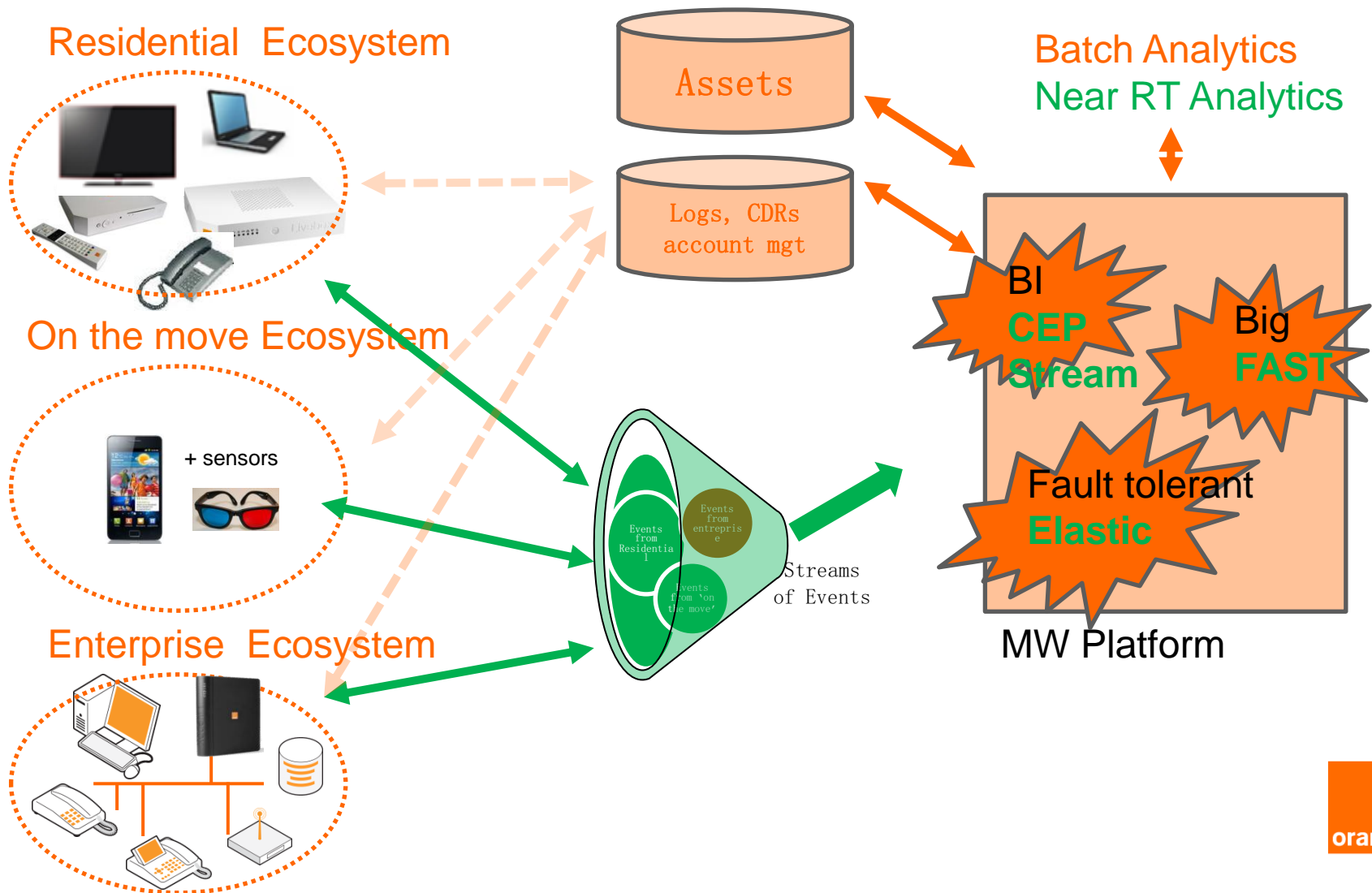





## Agenda

- Background
- Why Big Data?
- **Why Fast Data?**
- Why Big + Fast ?
- Key issues
- Use Cases
- Work in progress
- Conclusion
- References

# Why Fast data?







## Why Fast data? – Challenges

- "Real-Time" feature has a high potential value
- traditional Business Intelligence (BI) → deferred analysis
- emerging technologies dealing with Real-Time BI
  - mine huge amounts of structured & unstructured data streams,
  - correlate complex events in RT,
  - improve business & operation processes & proactivity.
- no "turn-key" product in the market today → integration is needed



## Why Fast data? – Use cases

### Promising Use cases

- churn prevention (critical situation detection, customer immersive marketing),
- customised charging and policy control, RT differentiated QoS,
- improved Product Launch:
  - Improved confidence of end users to take up new products
  - Faster product revenue achievements
- improved Customer Information:
  - Real-time information on customer behaviour
  - Optimisation of marketing & sales strategy




## Agenda

- Background
- Why Big Data?
- Why Fast Data?
- **Why Big + Fast ?**
- Key issues
- Use Cases
- Work in progress
- Conclusion
- References

## Why Big + Fast

- I am trying to join a colleague by phone many times with no success.


The Platform detects that he was just twitting 1  before. I receive a SMS recommendation telling me to contact him through

**Residential → Fast**

- In the last 3 minutes,  report problems of disconnection from the ADSL. The Platform detects the situation and proactively reports an added value Event to the Help Desk.

**Enterprise → Fast + Big**

- I am in front of the Orange Shop. My nephew bought 3 months  for a

 . His birthday is tomorrow. The Platform detects the situation  and

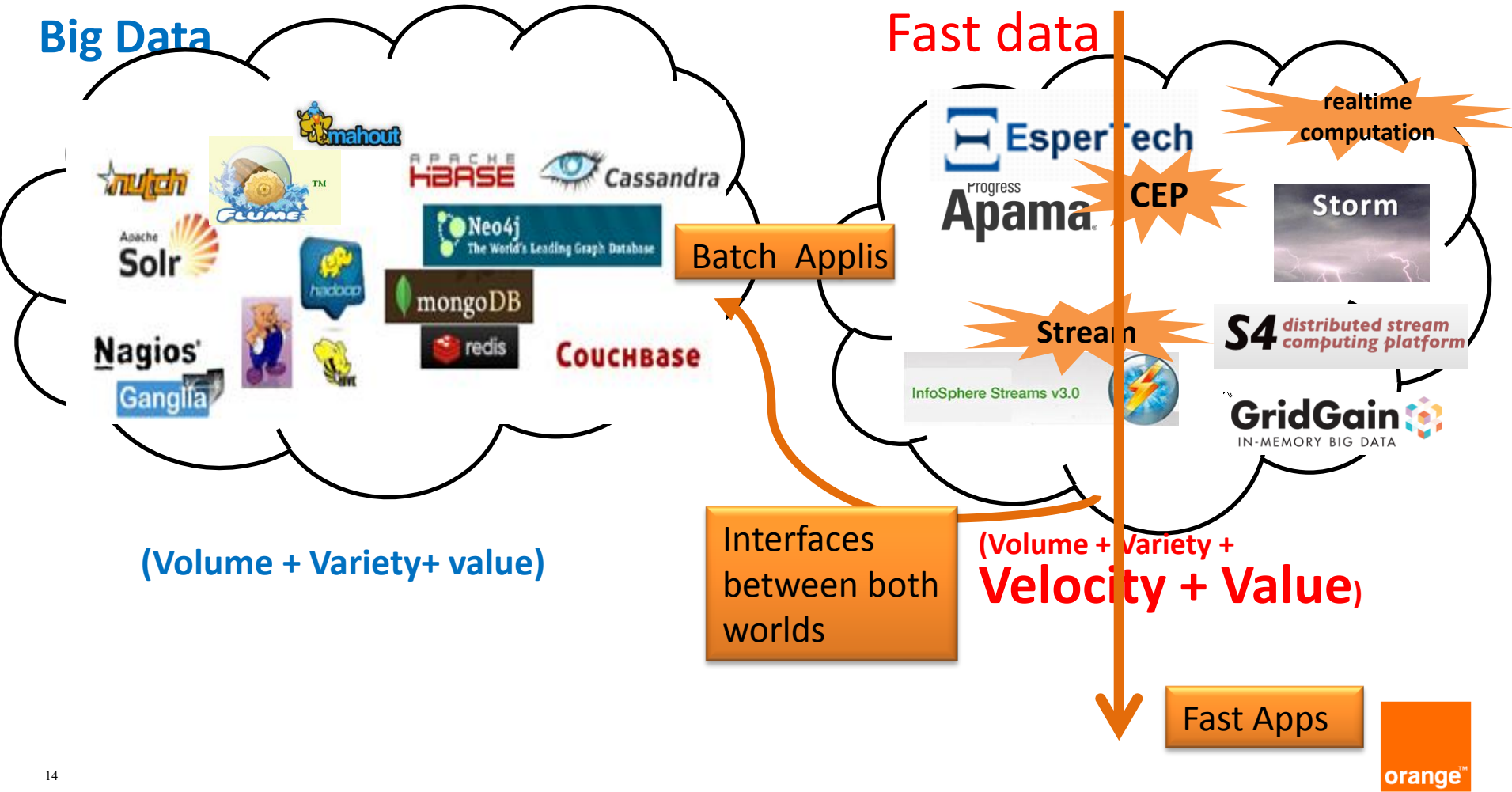
and recommends me to buy a gift for him



## Agenda

- Background
- Why Big Data?
- Why Fast Data?
- Why Big + Fast ?
- **Key issues**
- Use Cases
- Work in progress
- Conclusion
- References

# Key issues - Big + Fast Ecosystem



## Key issues - Input

### PLAY

- deliver an elastic and reliable architecture for dynamic and complex, event-driven interaction in large highly distributed and heterogeneous service systems
- FP7 – STREP project (October 2010 → October 2013)
- <http://www.play-project.eu/>

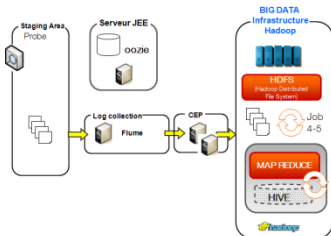
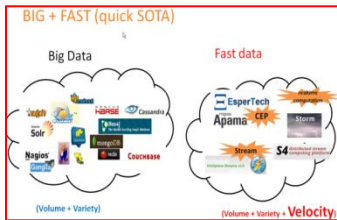
### SocEDA

- deliver an elastic and reliable architecture for dynamic and complex, event-driven interaction in large highly distributed and heterogeneous service systems + design tools
- ANR project ( November 2010 → November 2013 )
- <https://research.linagora.com/display/soceda/SocEDA+Overview>

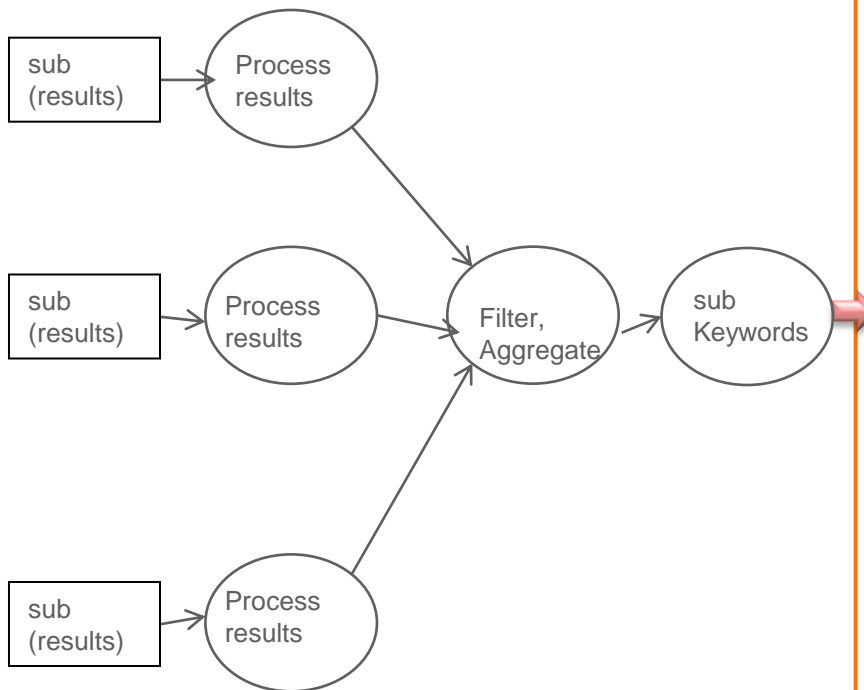


# Key issues - keywords

## Input



## Processing



## Keywords

- Stream Processing and CEP
- **CEP** with good **EPL**
- Scalability, Fault tolerance
- Hadoop Ecosystem
- **Publish and Subscribe**
- **Fast Middleware**
- Join Fast and Big Requests
- Big + Fast Architecture
- Performances



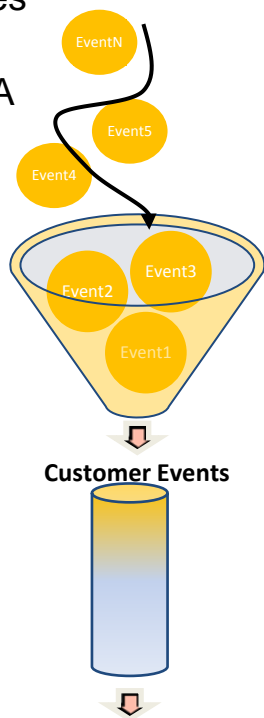


## Agenda

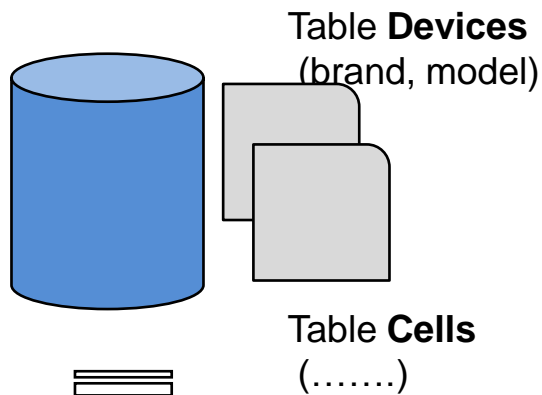
- Background
- Why Big Data?
- Why Fast Data?
- Why Big + Fast ?
- Key issues
- **Use Cases**
- Work in progress
- Conclusion
- References

# Use Cases - CEM

Events from Probes  
Stream of **Fast DATA**



**Big DATA**

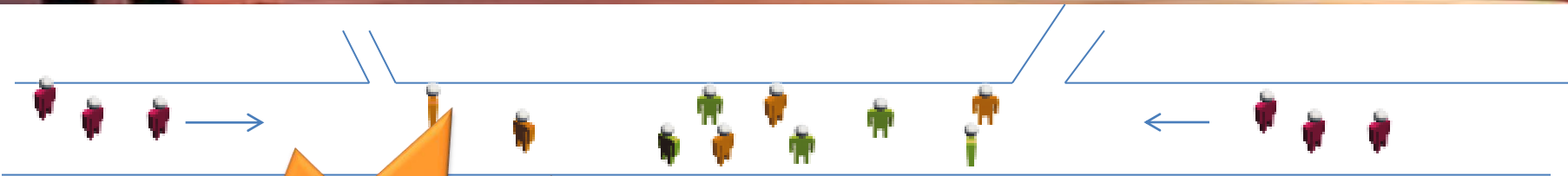





**Help desk**



- Rules**
- Identify a wrong configuration for a service
  - identify a missing event on an event stream
  - identify an error related to the terminal

## Use Cases - Promotional offers



-  **Near RealTime data:** they just arrived near the store
-  **Batch data:** they are there since a while but no Geolocation from them in the short past
-  **Future data:** they are going in the direction of the store

**Query 1**  
Which Orange Customers are able to **definitively** receive a visual message from the shop window?  
→ **Near RealTime** data

**Query 2**  
Which Orange Customers are able to enter the shop within 30 sec?  
→ **Batch data** + **Near RealTime** data

**Query 3**  
Which Orange Customers are able to enter the shop within 3 min?  
→ **Batch data** + **Near RealTime** data + **Future data**

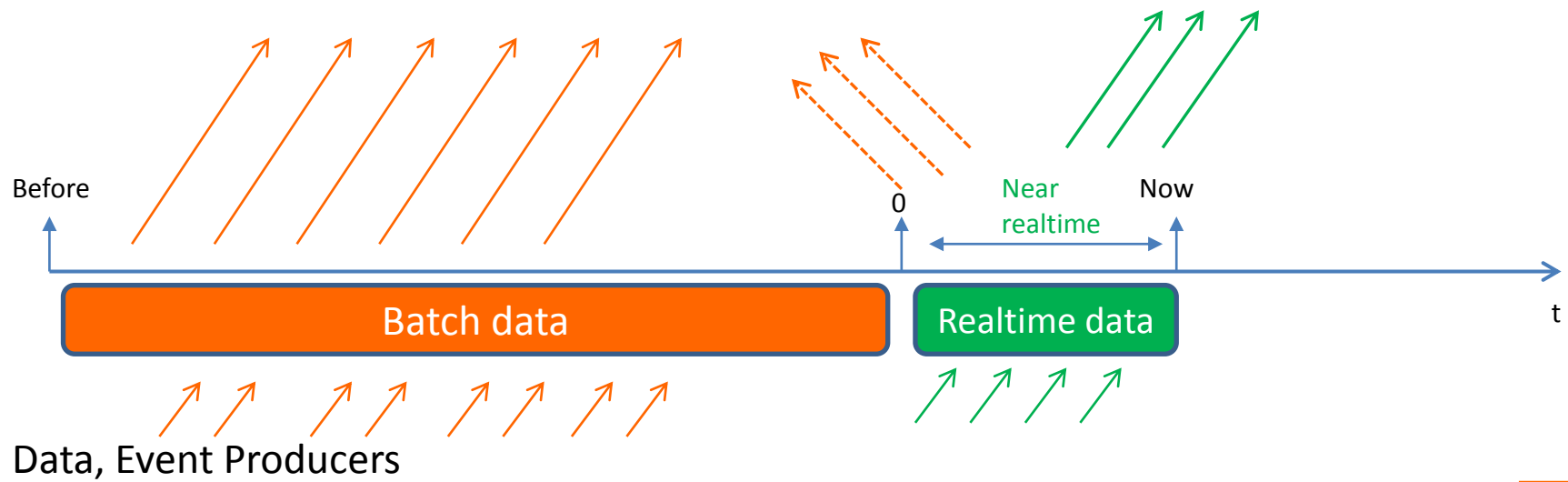


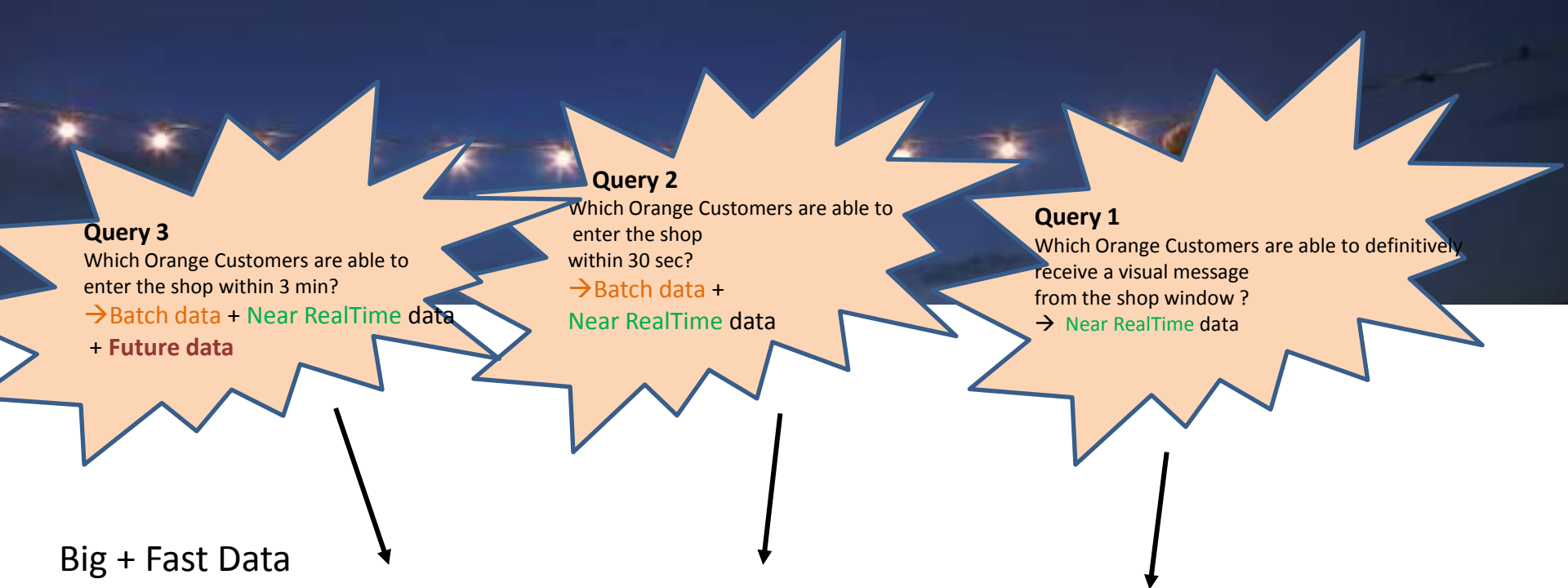
Classification



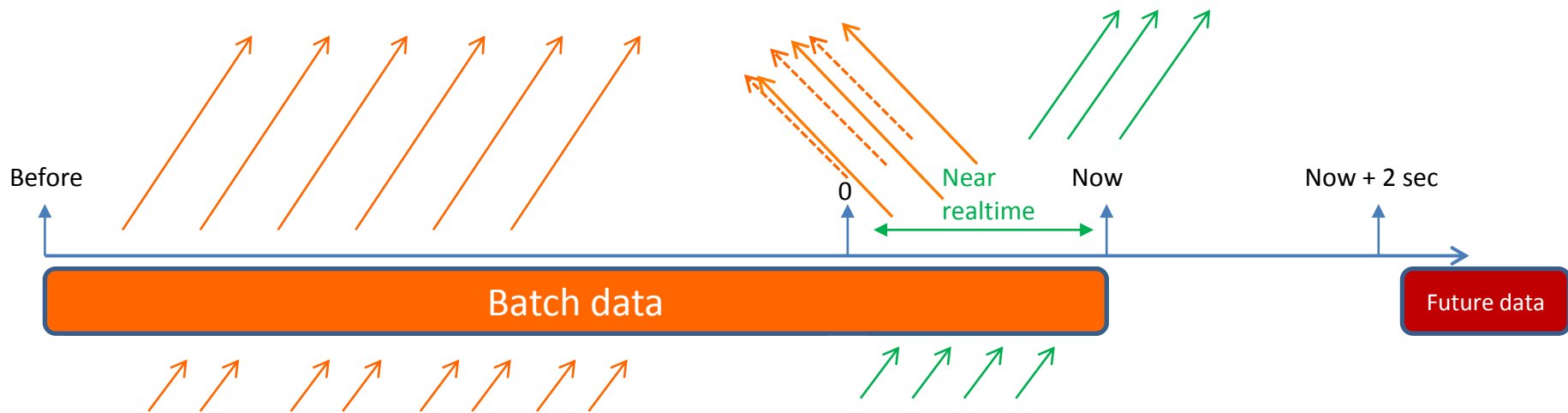


## Big + Fast Data





Big + Fast Data



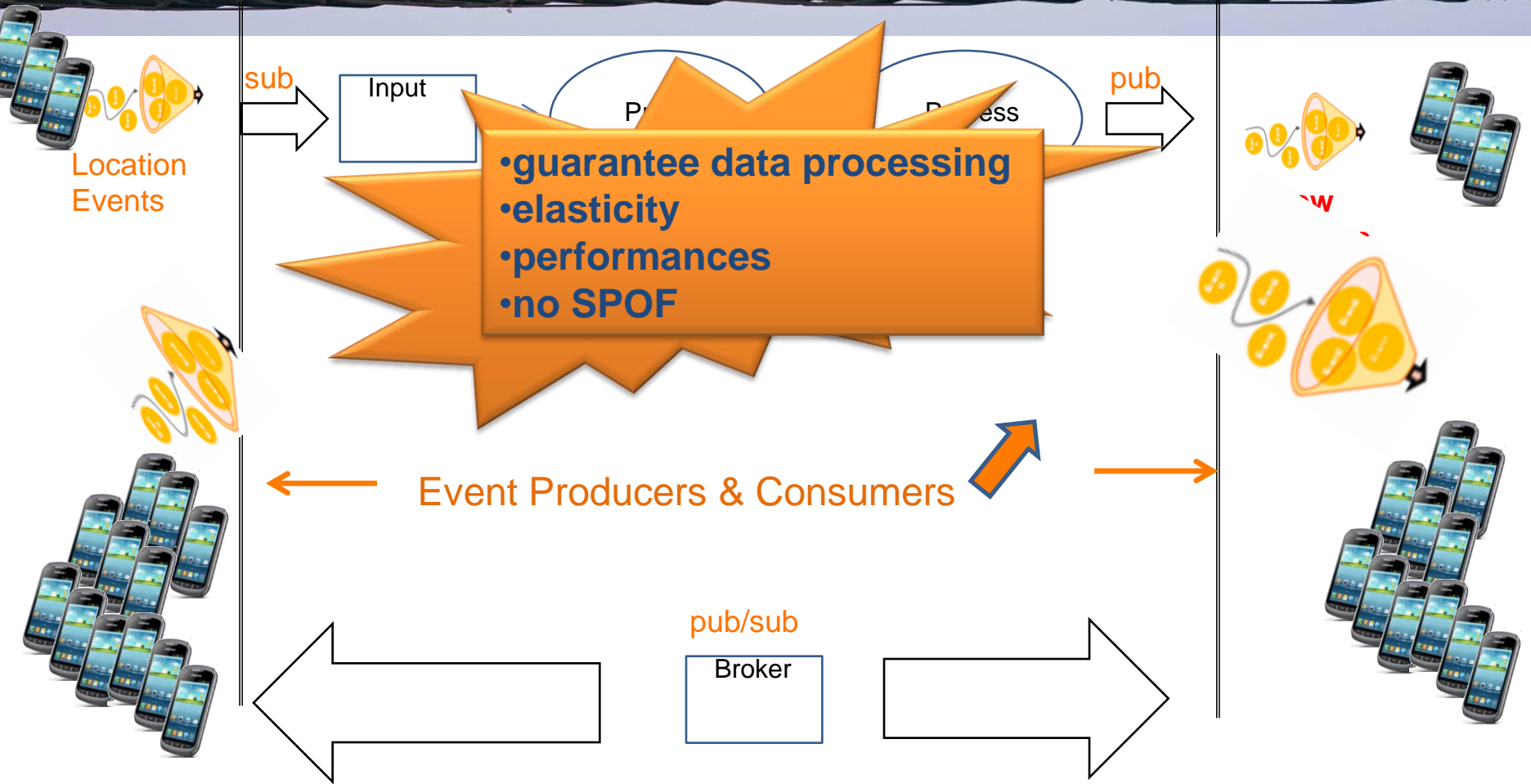
Data, Event Producers



## Agenda

- Background
- Why Big Data?
- Why Fast Data?
- Why Big + Fast ?
- Key issues
- Use Cases
- **Work in progress**
- Conclusion
- References

Work in progress



# Technologies

- CEP technology
  - high throughput – between 1,000 to 100k messages per second
  - low latency – react in real-time to conditions (from Msec to secs)
  - complex computations – detect patterns among events (event correlation), filter events, aggregate time or length windows of events, join event streams, trigger based on absence of events etc.
- reactive MW
  - fault tolerant, elastic, scalable
  - able to integrate technologies, language agnostic
  - open source



## Work in progress

- challenge platforms
  - PLAY, Soceda
  - Event oriented MW ( Storm , Esper , pub/sub MOM )
  - Other Big + Fast MW (revised SOTAs)
- next steps
  - describe functional and non functional requirements
  - implement relevant Use Cases
  - assess chosen platform (performances, scalability)



## Agenda

- Background
- Why Big Data?
- Why Fast Data?
- Why Big + Fast ?
- Key issues
- Use Cases
- Work in progress
- **Conclusion**
- References



## Conclusion

- actors (residential & enterprise market ), projects are constrained by a quick and accurate answer
- to develop these RT Added Value Apps
  - a **Big Data** Approach is mandatory
  - with traditional BI tools
  - but also
    - with special **Fast Data** MW adapted to RT or Near RT constraints
    - with BI tools adapted to these constraints
- some promising technologies, projects (open source oriented) are progressing on that 😊

## References

- Play Project <http://www.play-project.eu/>
- Soceda Project <https://research.linagora.com/display/soceda>
- Orange Labs <http://www.primezone.orange-labs.com/>
- Storm <http://storm-project.net/>
- Esper <http://esper.codehaus.org/>