

# Future technologies for efficient and agile handling of non-event data

Gianluca Cerminara  
Giovanni Franzoni - Andreas Pfeiffer (CERN EP-CMG-CO)  
Giacomo Govi (FNAL)  
Andrea Formica (IRFU CEA/Saclay)

CERN -  
EP Software R&D First Lightning Talks Session  
5-Feb-2018

# Non-event/Condition data?

- producers: detector/physics-object/field experts
  - e.g calibrations, alignment, detector geometries, Jet-Energy-Corrections, ML models...
- consumers: heterogeneous **zoo of applications**
  - need to know conditions @ certain point in time
    - **online** processing (*e.g. high level trigger - online data quality monitoring*)
    - offline **reconstruction**
    - data **analysis**
- content:
  - payload: arbitrary complex sets of values (payloads)
    - huge dynamic range in size
  - Interval-of-Validity (IOV): from a few seconds to years

# Use cases to be addressed

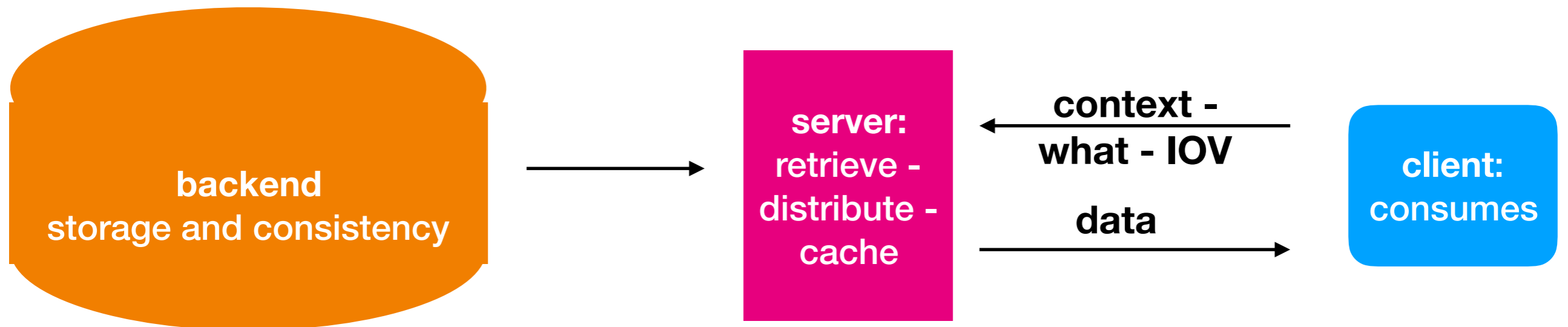
- Support **arbitrarily complex objects**
- Support access by different facilities
  - from big farm/cloud/HPC to analyzer's laptop
- **decouple back-end technology from interfaces** for consumption
  - > 1 backend solution for the same application
    - depending on use case (data-preservation, network-less setup for analysis or cloud usage) can use different back-end (ORACLE, other SQL, noSQL, DB+shared file-system)
- Reliable solutions for distribution (and caching)
  - built-in mechanism for **traceability** and **reproducibility**
  - support for quasi real-time consumers (e.g. HLT)
  - today's biggest challenge for deployment of quickly changing conditions (online calibrations)

# Another use case: Analysis

- **Analyses consume conditions** in RunII:
  - calibrations of high-level objects, ML models...
  - need to trace changes across the dataset (different LHC/detector conditions)
- HL-LHC rates → reprocessing difficult/impossible → need keep flexibility @ end of the processing chain
- Analyses use standalone/**framework-agnostic applications**
  - need to support modern analysis technologies (e.g. swan, hadoop...)
  - need to support python-based analysis specific frameworks
    - mainstream for ML applications
- The challenge: provide reliable and traceable access to conditions
  - without infrastructure for condition access
  - (→ need to be fast → analysis is IO bound)

# The seed for the technology R&D

- Atlas-CMS group started working on potential solution
  - prototype already tested on some of the use-cases
- A light-weight **client** requires:
  - move the intelligence to a **server**
  - data served via a commonly available medium (e.g http)
  - data structures to be portable (e.g JSON)



## meta-data

- "context"
- IOV

## data

- payloads

# Scope of the project

- **Future experiments**: generic and flexible framework with lightweight/no client
  - turnkey solution
  - allows focusing on physics studies
- Synergy with **LHC experiments and HL-LHC upgrades**
  - conditions more and more critical for high-level trigger selection
  - conditions allow to exploit multi-year datasets
  - get ready for transition out of ORACLE (if needed)
    - discussion and prototyping started in ATLAS-CMS common effort
    - other experiments joined the discussion and started implementing a model inspired by these lines (e.g BELLE-II)