21st International Conference on Computing in High Energy and Nuclear Physics **CHEP2015** Okinawa Japan: April 13 - 17, 2015

WLCG
Worldwide LHC Computing Grid

Evolution of the ALICE computing model in Run 3

Tatsuya Chujo
(for the ALICE collaboration)

WLCG workshop
April 12, 2015, OIST, Japan

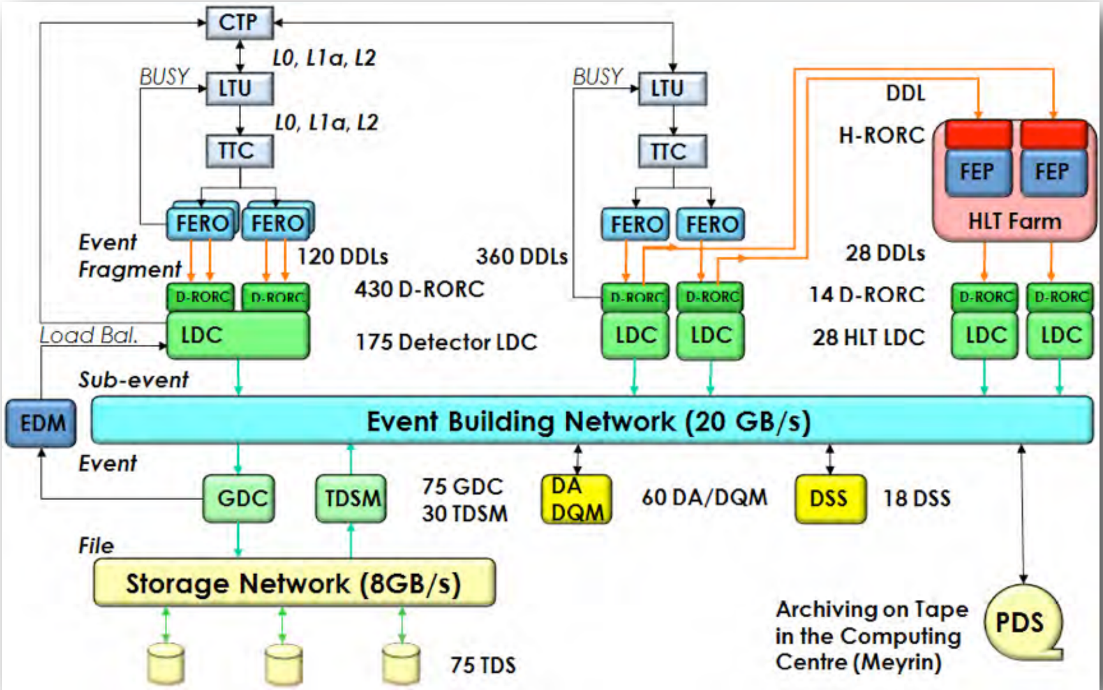
筑波大学
University of Tsukuba

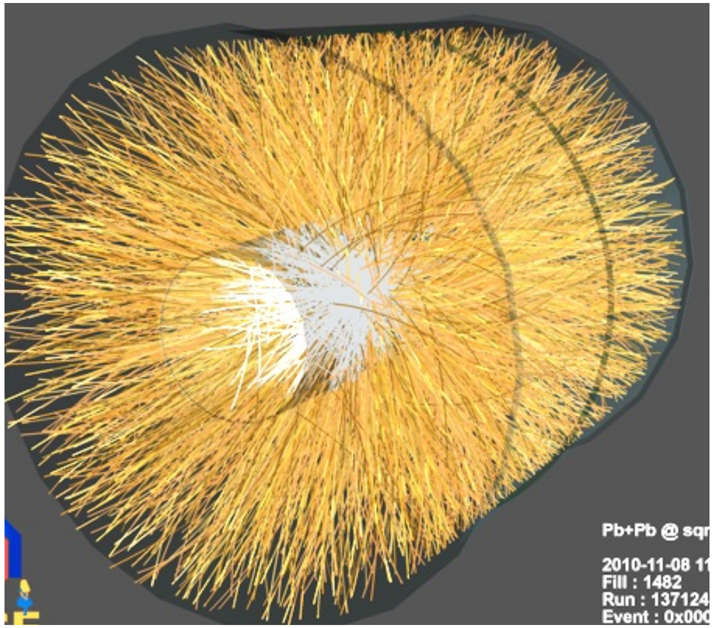
ALICE

ALICE data collection (current)

Nominal LHC beam crossing at 40 MHz

ALICE:
Multi-level trigger system needed:
40 MHz → a few kHz

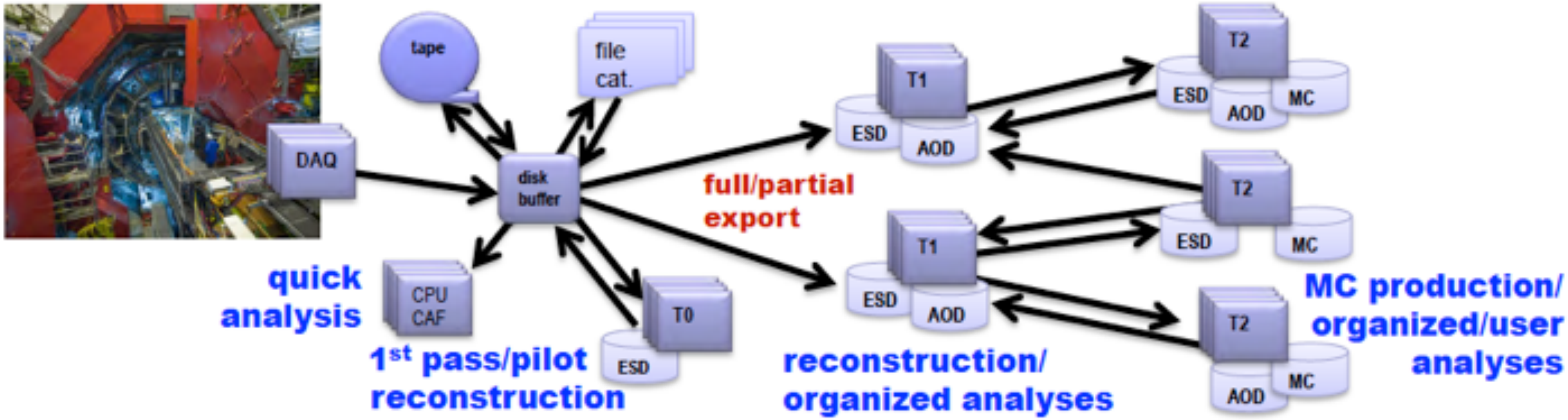




Single Pb-Pb collision events
($\sqrt{s_{NN}}$ = 2.76 TeV)

<u>Online:</u>
1) Reject background
2) Select most interesting interactions
3) Custom computer to reduce the total data volume

Computing model (Run1 and Run2, -2018)



slide by T. Sugitate

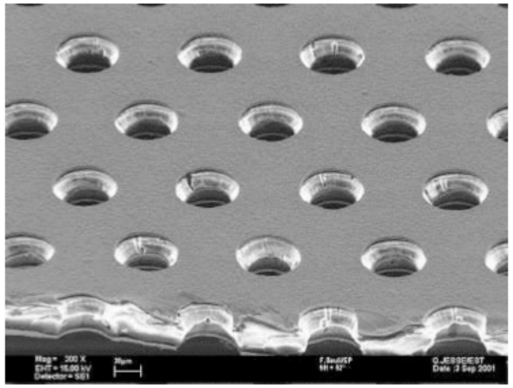
ALICE upgrade (2018-)

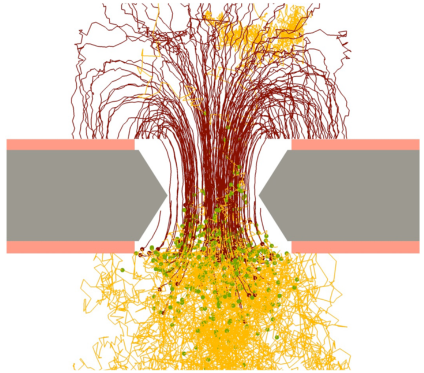
<u>ALICE upgrade; high rate capability</u>

GEM-TPC continuos high rate readout

ITS Silicon high rate readout

DAQ (RCU etc.)



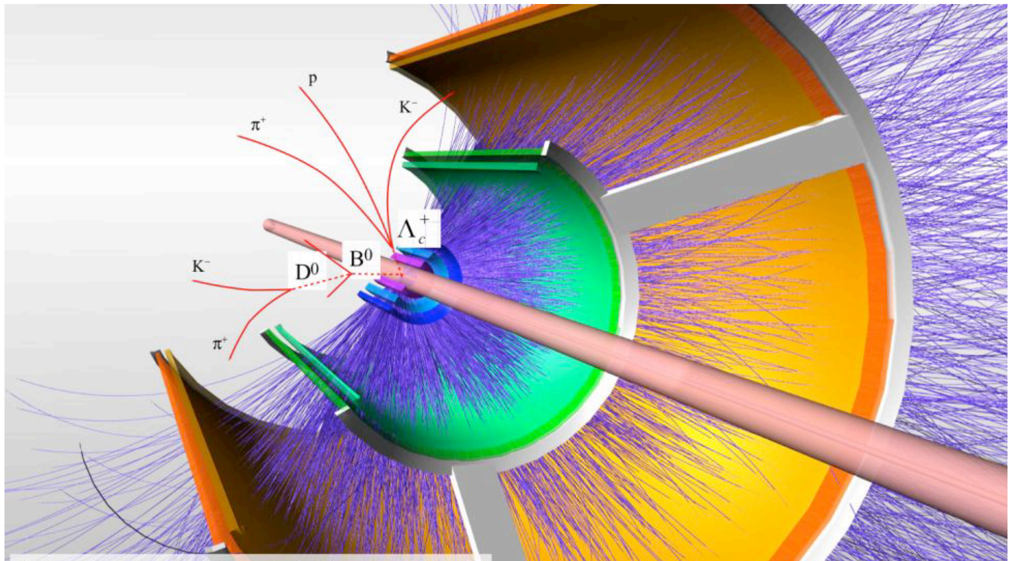


Standard GEM
Pitch=140µm
Hole ϕ=70µm

For LHC high luminosity upgrade, Pb-Pb @50kHz

Record all MB events, x100 statistics
（Unique capability in ALICE）

→Access to high precision measurements and rare probes



Sketch of the ITS upgrade with half a Pb-Pb event superimposed

Physics Goals:

Measure

- heavy quarks, photons, lepton pairs
azimuthal anisotropy
- Jet w/ PID hadron simultaneously

ALICE Upgrade Plans: (2018-)

- **Current: reducing the event rate from 40 MHz to ~ 1 kHz**
  - Select the most interesting particle interactions
  - Reduce the data volume to a manageable size
- **After 2018:**
  - Much more data (**X 100**) because:
    - Higher interaction rate
    - More violent collisions → More particles → More data (1 TB/s)
    - Physics topics require measurements characterized by;
      - Very small signal/background ratio → large statistics
      - Large background → traditional triggering or filtering techniques very inefficient for most physics channels
    - Read out all particle interactions (PbPb) at the anticipated interaction rate of **50 kHz**
    - **No more data selection**
      - Continuous detector read-out
      - Read-out and process all interactions with a standard computer farm.
      - ~1,500 nodes with the computing power expected by then

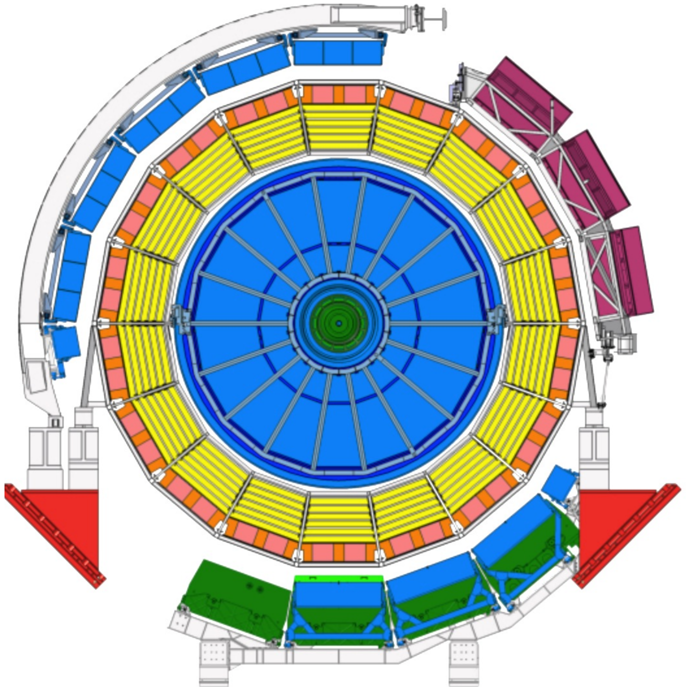
➡ **Total data throughput out of the detectors: 1 TB/s**

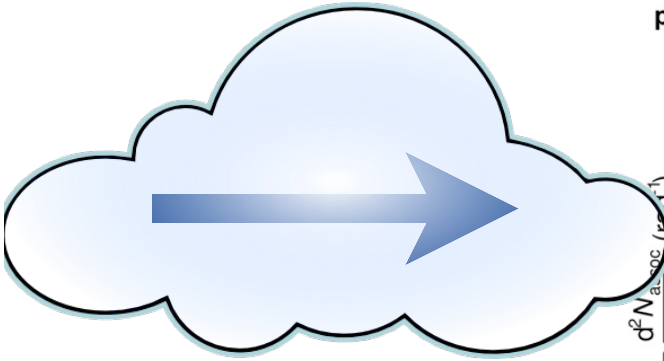
Expected data bandwidth (after 2018-)

Detector	Input to Online System (GB/s)	Peak Output to Local Data Storage (GB/s)	Average Output to Computing Center (GB/s)
TPC	1,000	50	8
TRD	81.5	10	1.6
ITS	40	10	1.6
Others	25	12.5	2
TOTAL	**1,146.5**	**82.5**	**13.2**

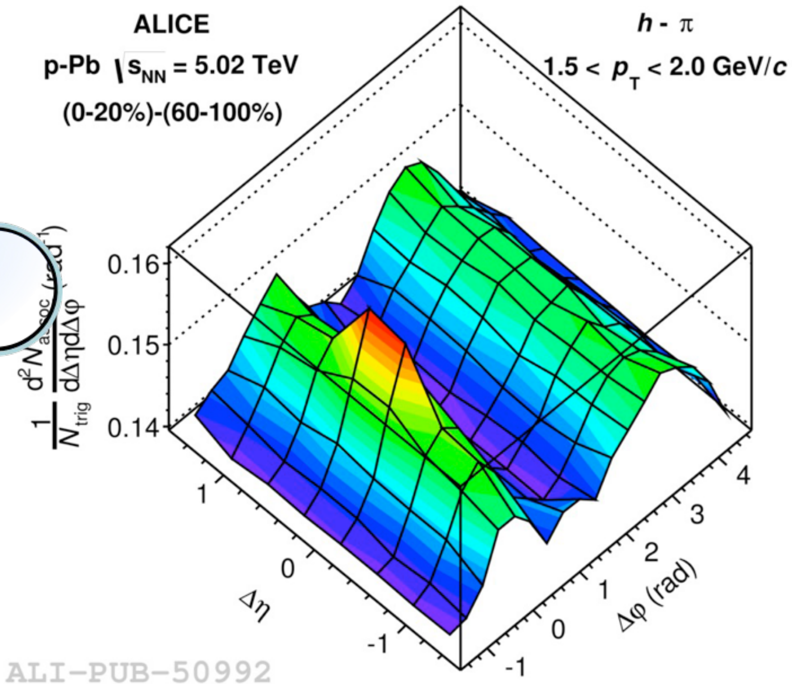
Note: LHC luminosity variation during fill and efficiency taken into account for average output to computing center

The ALICE Online-Offline (O2) Project





ALICE



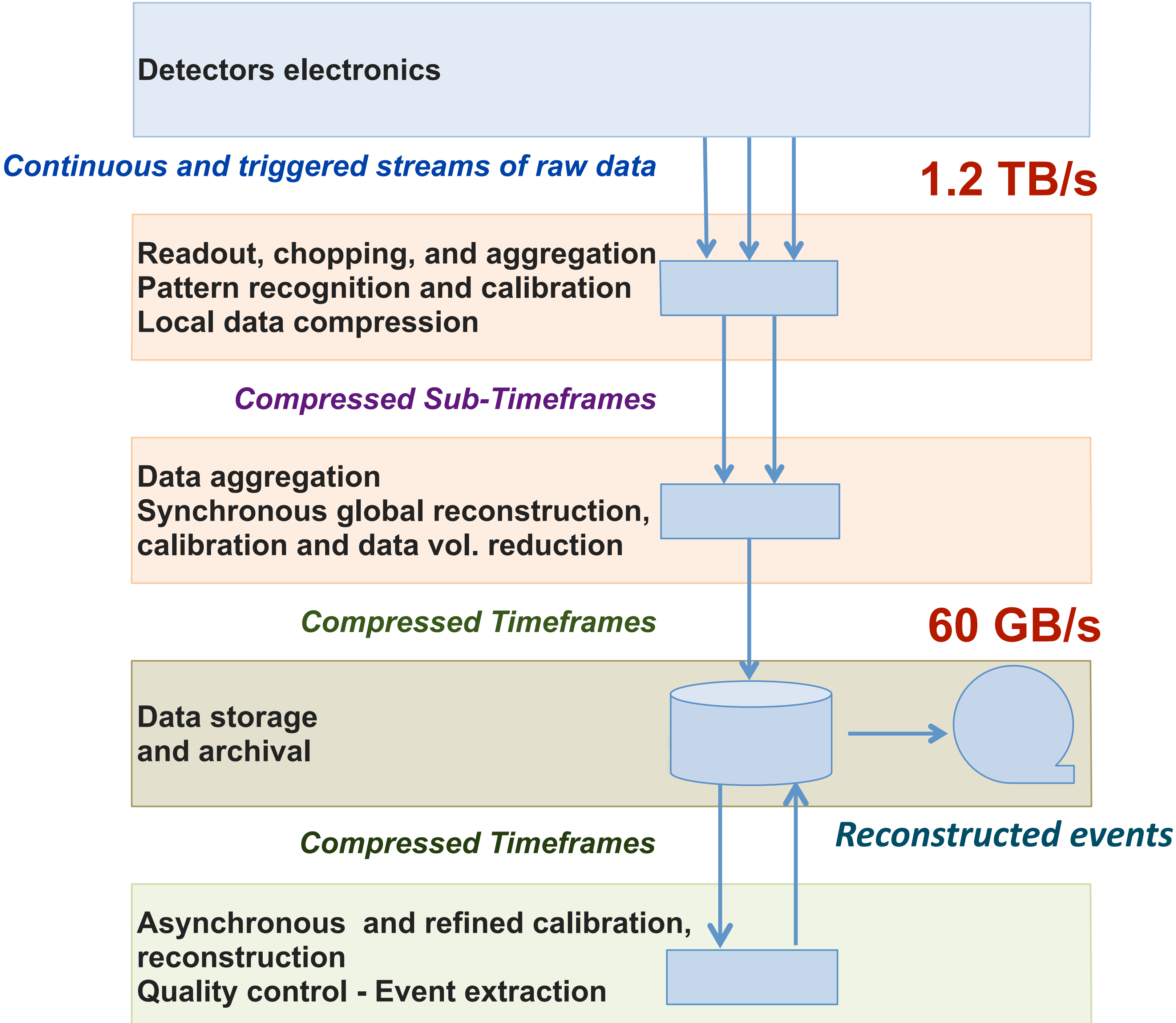
- From Detector Readout to Analysis:
- What is the "optimal" computing architecture?

- Handle **>1 T Byte /s** detector input
- Support for continuous readout
- Online reconstruction to reduce data volume
- Common hardware and software system developed by the DAQ, HLT, Offline teams

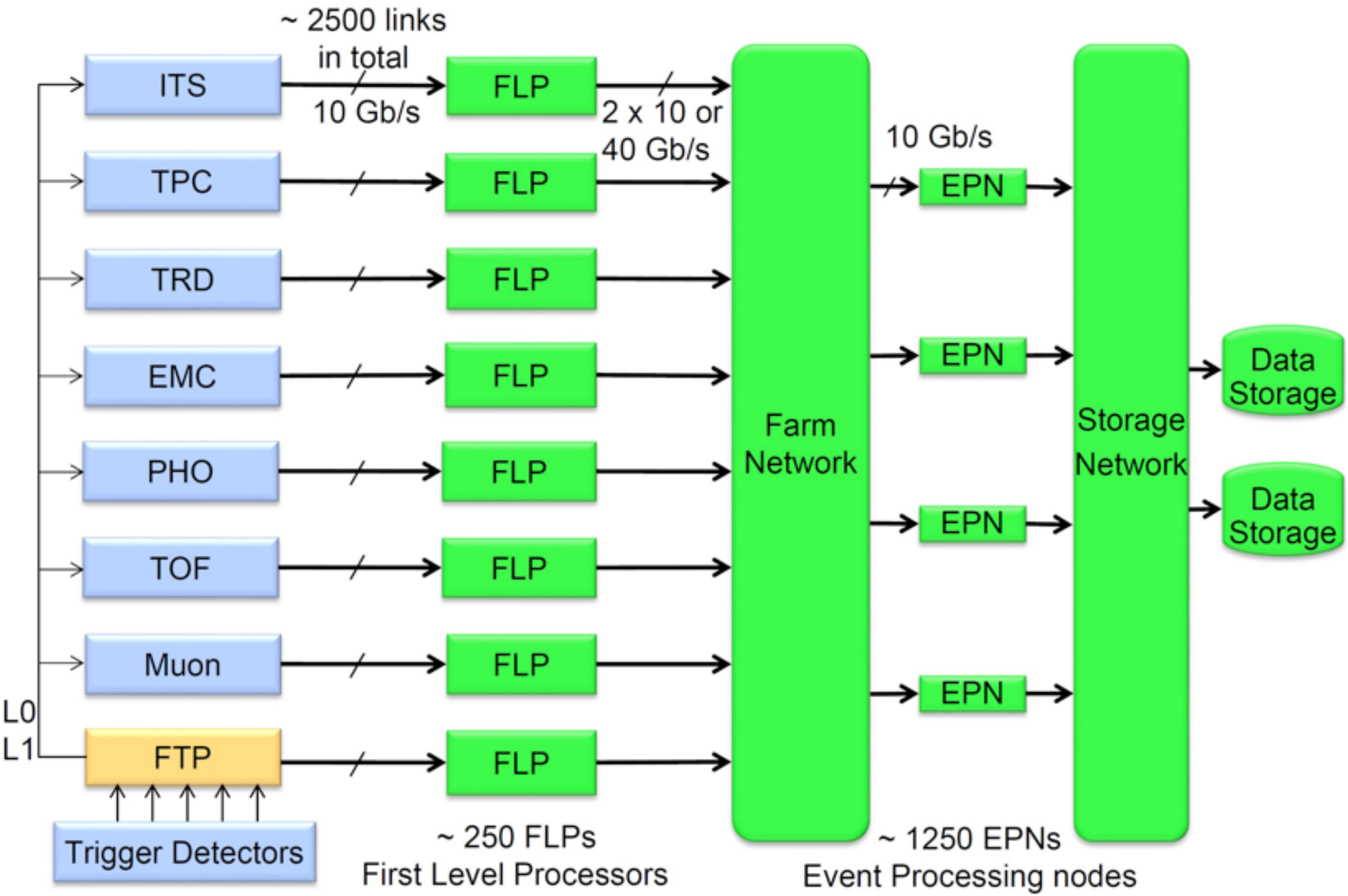
Functional Requirements of the O2 system

- ✓ Data fully compressed before data storage.
- ✓ Reconstruction with calibrations of better quality.
- ✓ Grid capacity will evolve much slower than the ALICE data volume.
- ✓ Data archival of reconstructed events of the current year to keep Grid networking and data storage within ALICE quota.
- ✓ Needs for local data storage higher than originally anticipated

Basic idea of the O2 system

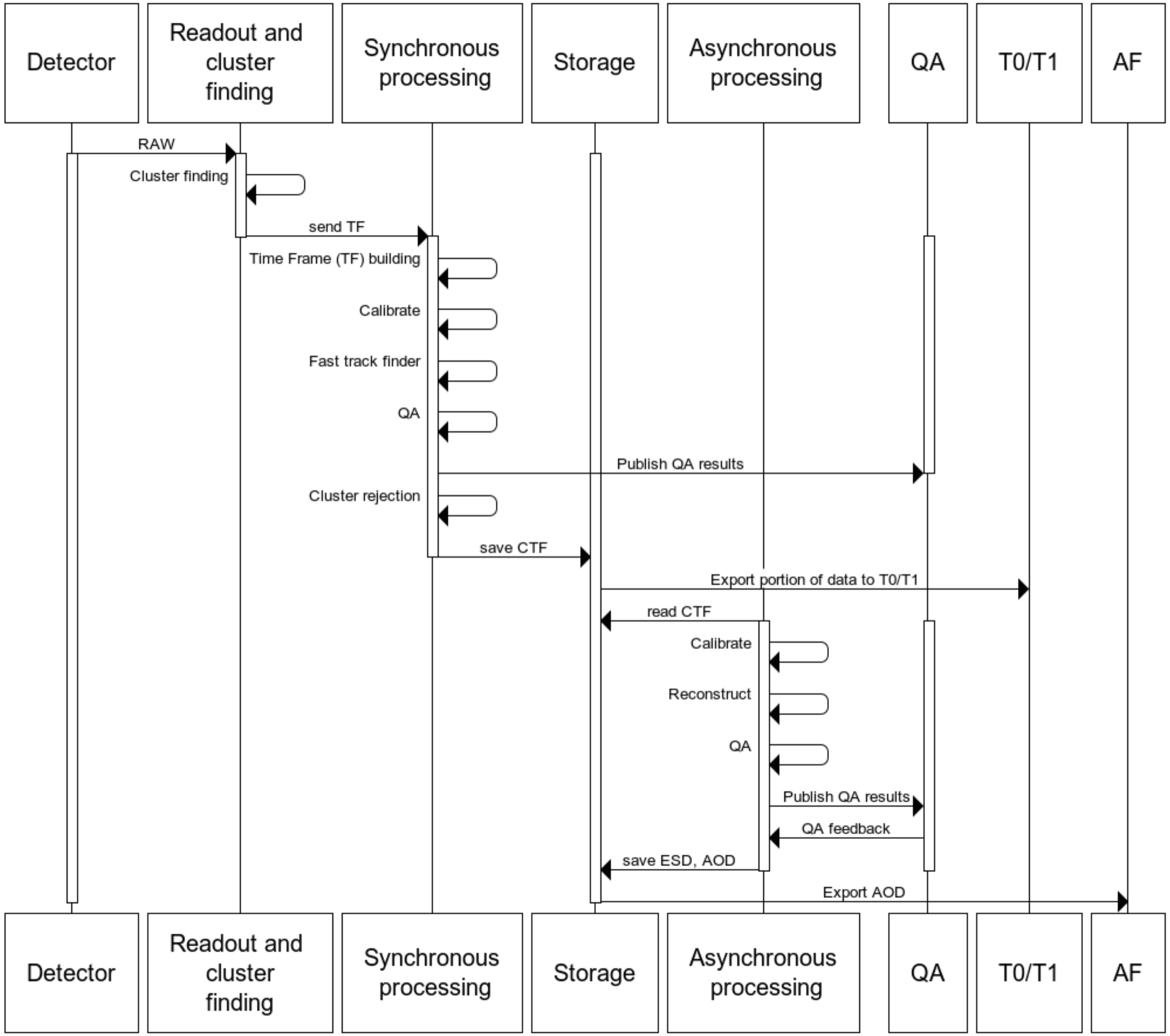


The ALICE O2 Hardware Architecture



Computing model (Data flow)

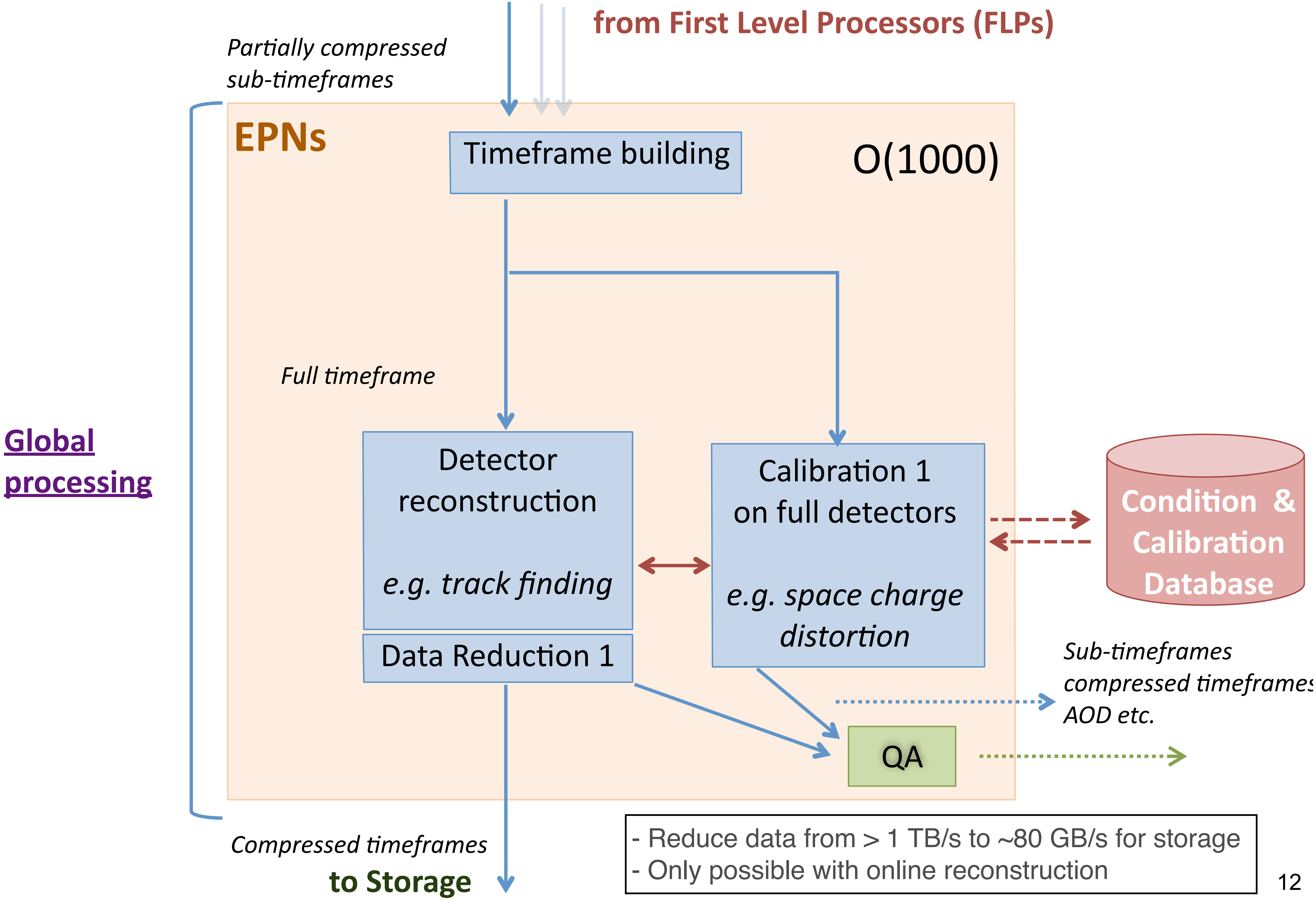
Acronym	Description
RAW	Raw data as it comes from the detector.
CTF	Compressed Time Frame containing the history of OM(100 ms) of detector readout information in the form of identified clusters that belong to identified tracks.
ESD	Event Summary Data.
AOD	Analysis Object Data for physics analysis.
HISTO	The subset of AOD information specific for a given analysis.
MC	Montecarlo simulation



by Pierre Vande Vyvre

The ALICE O2: Data Reduction (I)

by Pierre Vande Vyvre (modified)



- Reduce data from > 1 TB/s to ~80 GB/s for storage
- Only possible with online reconstruction

The ALICE O2: Data Reduction (II)

ALICE

	Dataflow Stage	Data Reduction Factor	Event Size (MByte)
	Raw Data	1	700
FEE →	Zero Suppression	35	20
High Level Trigger	Clustering & Compression	5 – 7	~ 3
High Level Trigger	Remove clusters not associated to relevant tracks	2	1.5
High Level Trigger	Data Format Optimization	2 – 3	< 1

slide by A. Uras (IC3INA 2013)

The ALICE O2: Data Reduction (III)

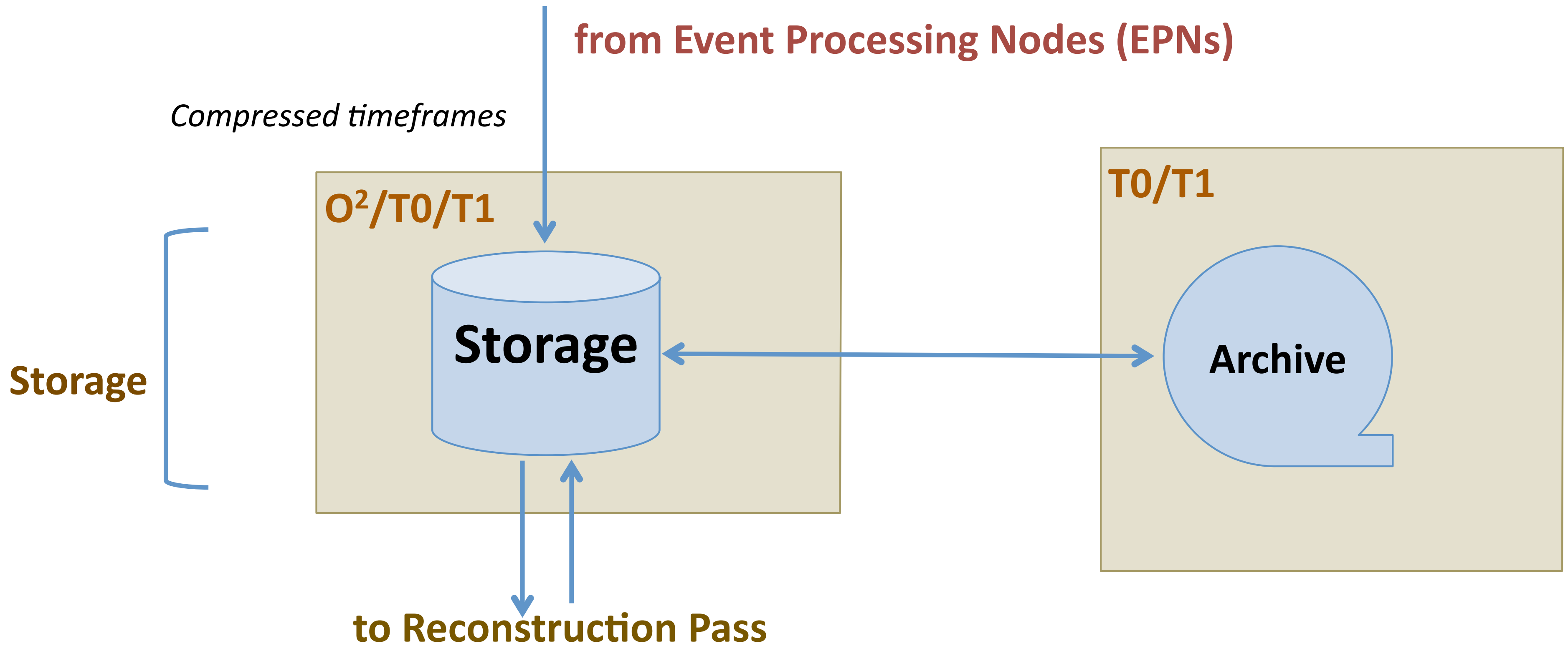
Detector	Event Size (MByte)	
	After Zero Suppression	After Data Compression
TPC	20.0	1.0
TRD	1.6	0.2
ITS	0.8	0.2
Others	0.5	0.25
TOTAL	**22.9**	**1.65**

- Data compression factors ranging from 2 to 20 according to the detector
- TPC still accounts for 60% of the total event size

slide by A. Uras (IC3INA 2013)

The ALICE O2: Data Storage

by Pierre Vande Vyvre (modified)

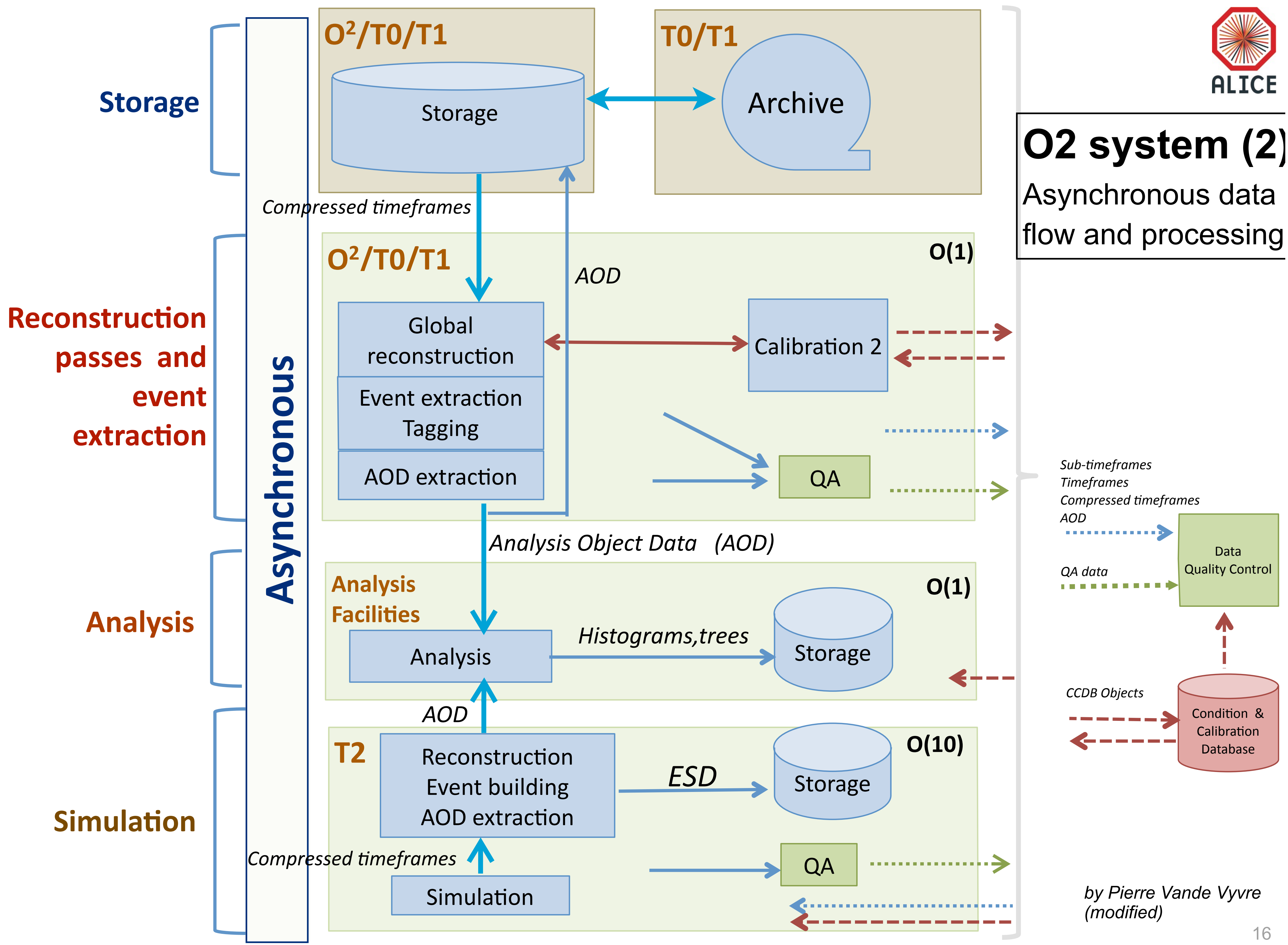


•Data in "intermediate" formats (not directly usable for physics analysis):

- 80 GByte/s peaks to be handled, distributed over ~1,250 nodes
- Average load of 15 GByte/s
- Local storage in O2 system
- Permanent storage in computing center

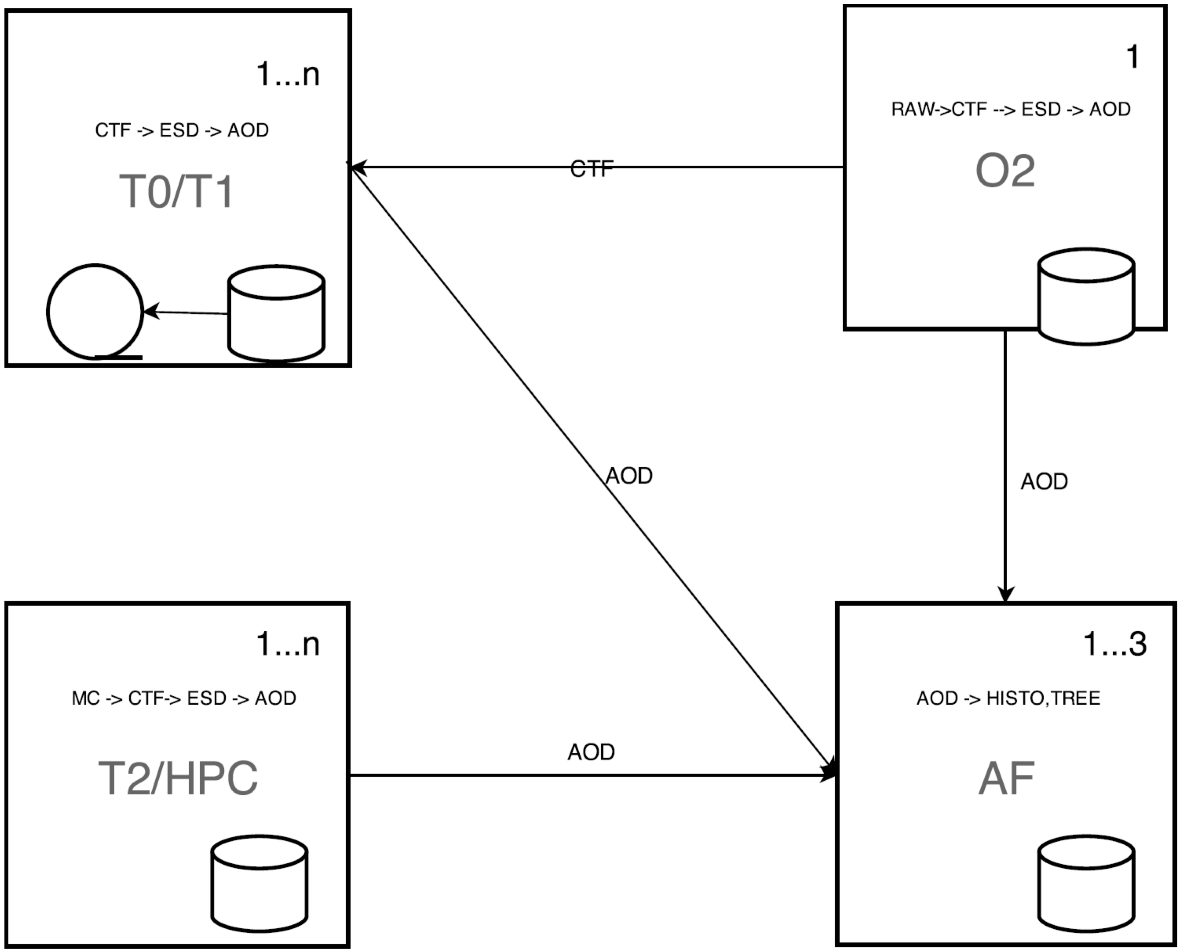
•Data in "final" formats (usable for physics analysis):

- GRID storage, accessible by experiment's users


O2 system (2)
Asynchronous data flow and processing
ALICE
Storage
Reconstruction passes and event extraction
Analysis
Simulation
Asynchronous
O²/T0/T1
Storage
T0/T1
Archive
Compressed timeframes
O²/T0/T1
O(1)
AOD
Global reconstruction
Event extraction Tagging
AOD extraction
Calibration 2
QA
Analysis Object Data (AOD)
Analysis Facilities
O(1)
Analysis
Histograms,trees
Storage
AOD
T2
O(10)
Reconstruction Event building AOD extraction
ESD
Storage
Compressed timeframes
Simulation
QA
Sub-timeframes
Timeframes
Compressed timeframes
AOD
QA data
Data Quality Control
CCDB Objects
Condition & Calibration Database
by Pierre Vande Vyvre (modified)

Computing model (O2 processing flow)

Facility	Function
O2	**ALICE Online-Offline Facility at LHC Point 2.** Online reconstruction during the run. Provides data storage capacity. After data taking: runs the calibration and reconstruction tasks.
T0	**CERN Computer Center** facility providing CPU, storage and archiving resources.
T1	Grid site connected to T0 with **high bandwidth network links (100+ Gb)** providing CPU, storage and archiving resources. Reconstruction and calibration tasks
T2	Regular grid site with good network connectivity (10+ Gb); running **simulation jobs**.
AF	**Dedicated Analysis Facility of HPC** type that collects and stores AODs produced elsewhere and runs the organized **analysis activity**.



- Maintain the advantages of the Grid and the analysis trains
- Make it more open and more effective

by Pierre Vande Vyvre

Impact on the network

- **2015-2017:**
  - LHC Run2 data taking, x2 more heavy ion data.
  - Data analysis on Run-2 (+Run-1)
  - Almost no change from Run-1 scheme. Due to data larger data volume, network traffic will increase, at least x2.
- **2018- beyond:**
  - LHC Run-3 data taking, x100 more data.
  - Architecture change (O2) applied.
    - O2, AF, and T0/T1/T2 scheme.
  - Significant data reductions, reconstruction in O2 mainly, and analysis → reduce data volume.

1. Can keep the similar network traffic as Run-2?
2. Or if we have **AF (using HPC)**, then it will need more network traffic than that in Run-2 → Accelerate local physics analysis.

Summary and Outlook

- **ALICE computing upgrades on online-offline for the data taking after 2018 is ongoing.**
  - Continuous minimum bias event readout at **50 kHz in Pb-Pb collisions**.
  - **1 TB/s raw data** from detector, need a significant data reduction down to 80 GB/s to storage, and make a physics outputs timely.
  - O2 Scheme: Online reconstruction and calibration by O2 (near ALICE) & T0/T1, organized analysis at Analysis Farm (AF), and simulation at T2.
- Designing based on physics requirements is completed.
- Intensive works on modeling, technologies (processing platform & network), O2 prototyping.
- Technical Design Report (TDR) is progressing. It will be submitted to LHCC in April 2015.