



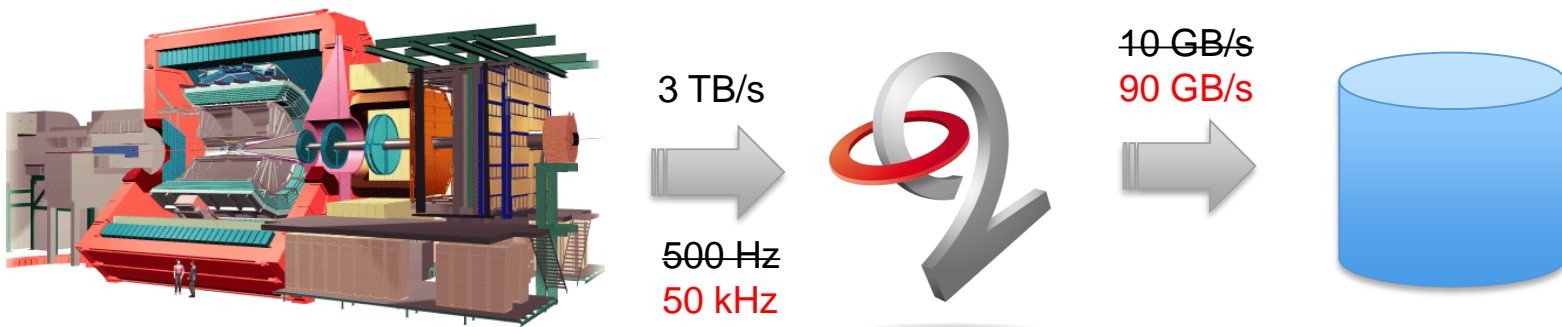
Analysis in Run 3

Predrag Buncic



Run 3 data taking objectives

- For Pb-Pb collisions:
 - Reach the target of 4 **13** nb⁻¹ integrated luminosity in Pb-Pb for rare triggers.
- The resulting data throughput from the detector has been estimated to be greater than 1TB/s for Pb–Pb events, roughly two orders of magnitude more than in Run 1

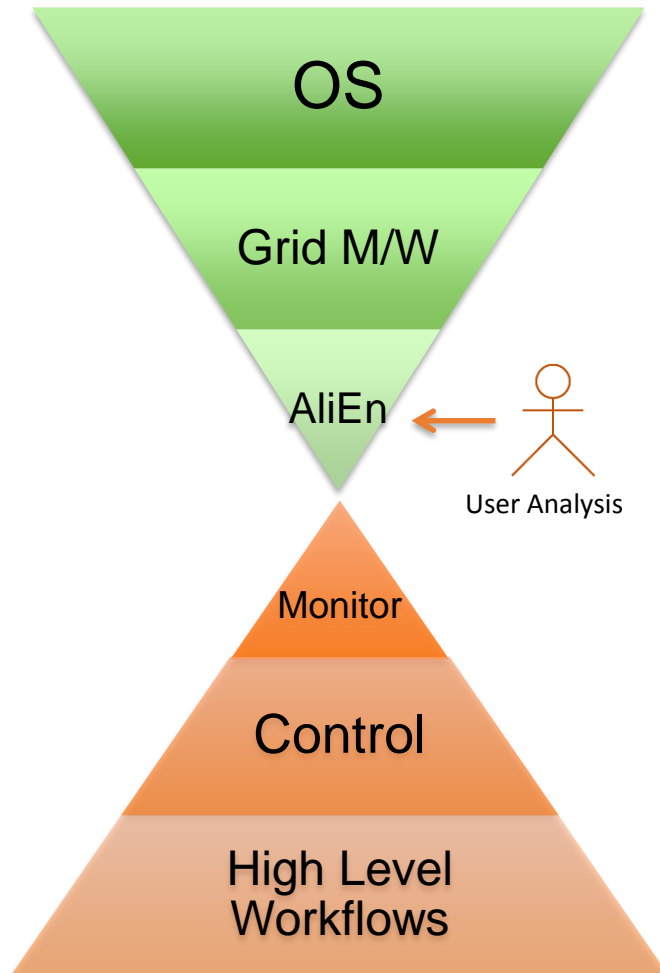


Computing Model in Run 1&2



- Computing model built on top of WLCG services
 - Any job can run anywhere and access data from any place
 - Scheduling takes care of optimizations and brings “computing to data”
 - Every file and its replicas are accounted in the file catalogue
- Worked (surprisingly) well during Run 2 and Run 2

Difficult to change...



- In production

- AliEn – ALICE Grid Environment
- Dates back to early days of the Grid (2000)
- Intensely developed and debugged for over 10 years
- Pioneered use of new concepts and technologies in Grid computing
 - Web services
 - Central Task Queue
 - Pilot jobs
 - Use of xrootd for data transport

- Under development

- Extensive monitoring
- jAliEn
- Complex production workflows for reconstruction and simulation
- Organized analysis

New in Run 3: O2 facility

- + 463 FPGAs
 - Detector readout and fast cluster finder
- + 100'000 CPU cores
 - To compress 1.1 TB/s data stream by overall factor 14
- + 3000 GPUs
 - To speed up the reconstruction
 - 3 CPU¹⁾ + 1 GPU²⁾ = 28 CPUs
- + 60 PB of disk
 - To buy us an extra time and allow more precise calibration

= Considerable (but heterogeneous) computing capacity that will be used for Online and Offline tasks

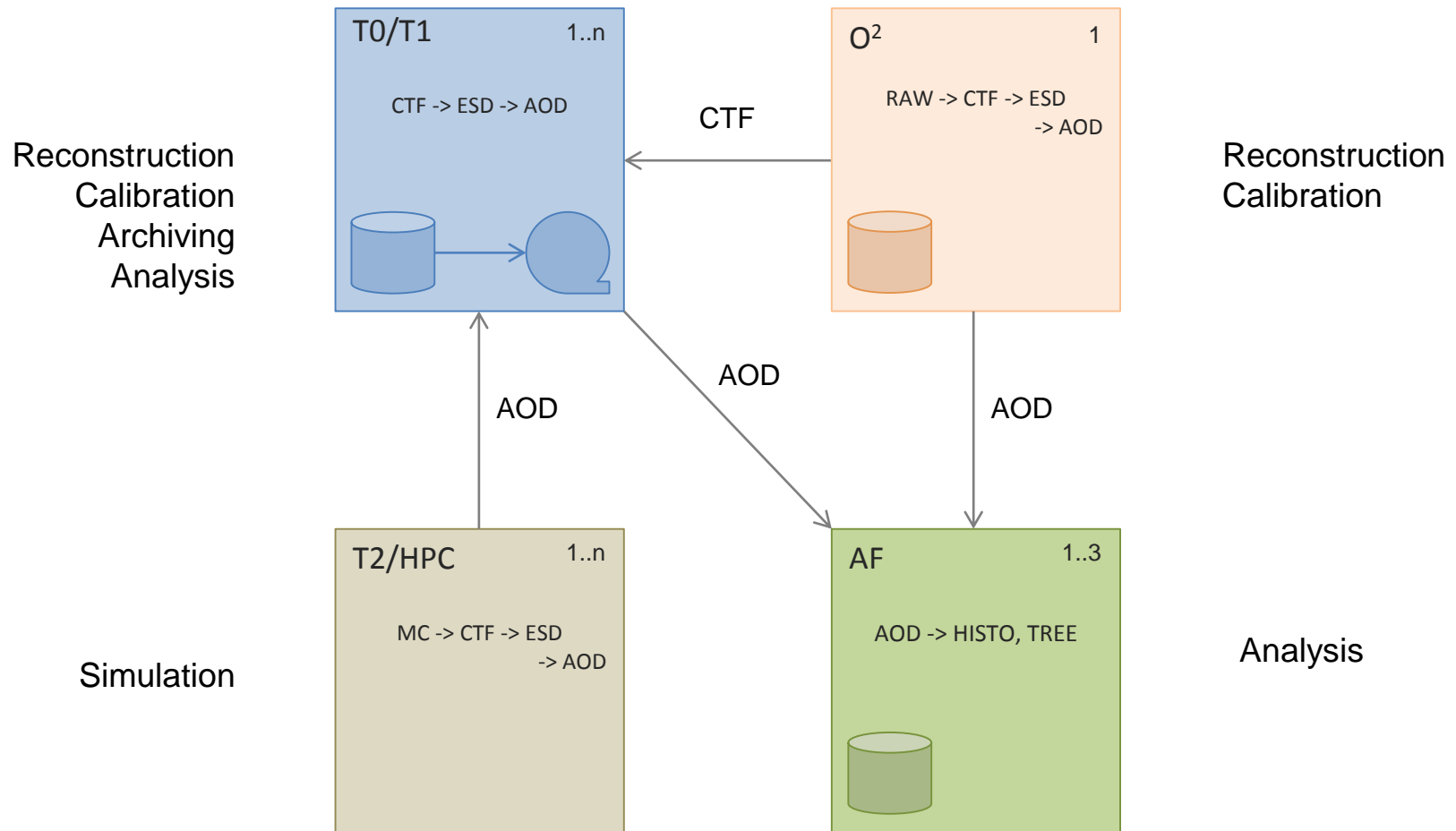
- ✧ Identical s/w should work in Online and Offline environments

¹⁾ Intel Sandy Bridge, 2GHz, 8 core, E5-2650

²⁾ AMD S9000

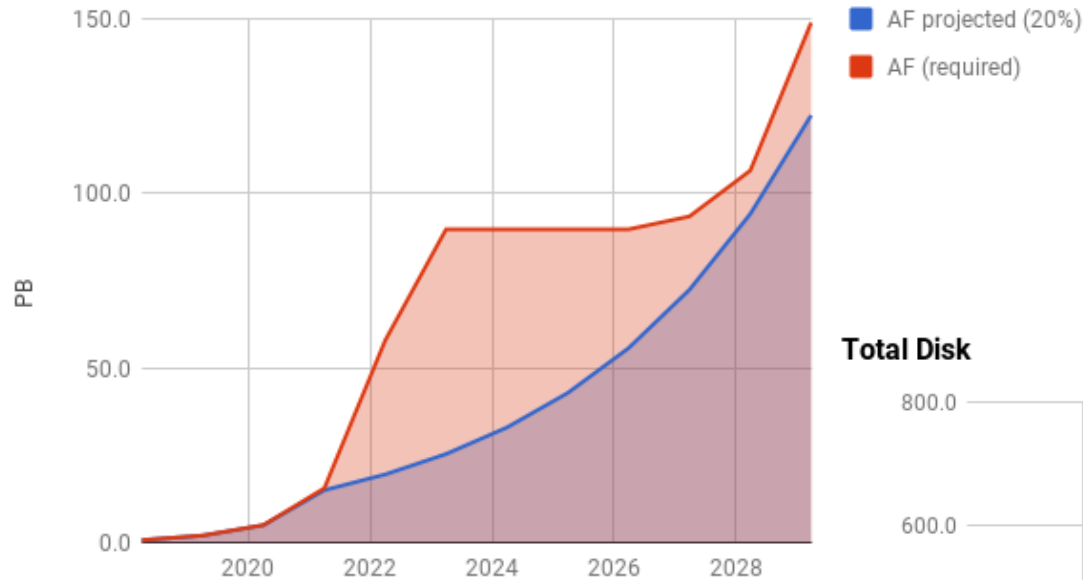


New (predominant) roles of Tiers in Run 3



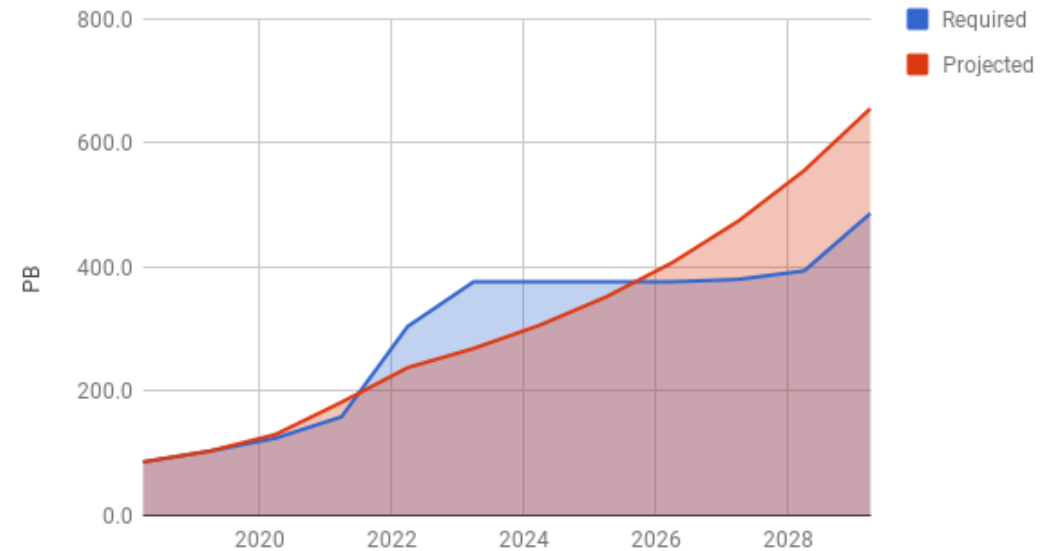
AOD data volume

Disk at AF



Assuming AOD size to be 20% of CTF

Total Disk



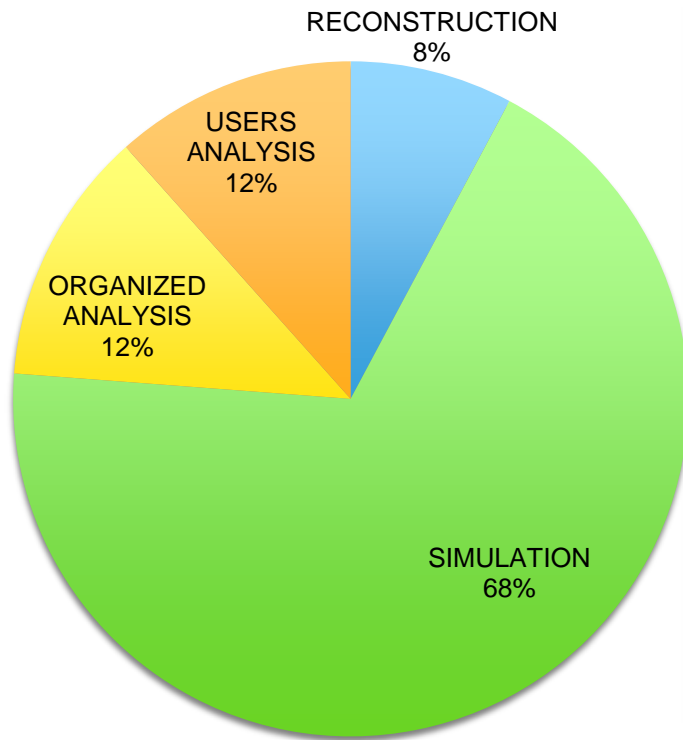
Changes to data management policies

- Only one instance of each raw data file (CTF) stored on disk with a backup on tape
 - In case of data loss, we will restore lost files from the tape
 - O2 disk buffer should be sufficient to accommodate CTF data from the entire period.
 - As soon as it is available, the CTF data will be archived to the Tier 0 tape buffer or moved to the Tier 1s
- All other intermediate data created at various processing stages is transient (removed after a given processing step) or temporary (with limited lifetime)
 - Only CTF and AODs are archived kept on disk to tape

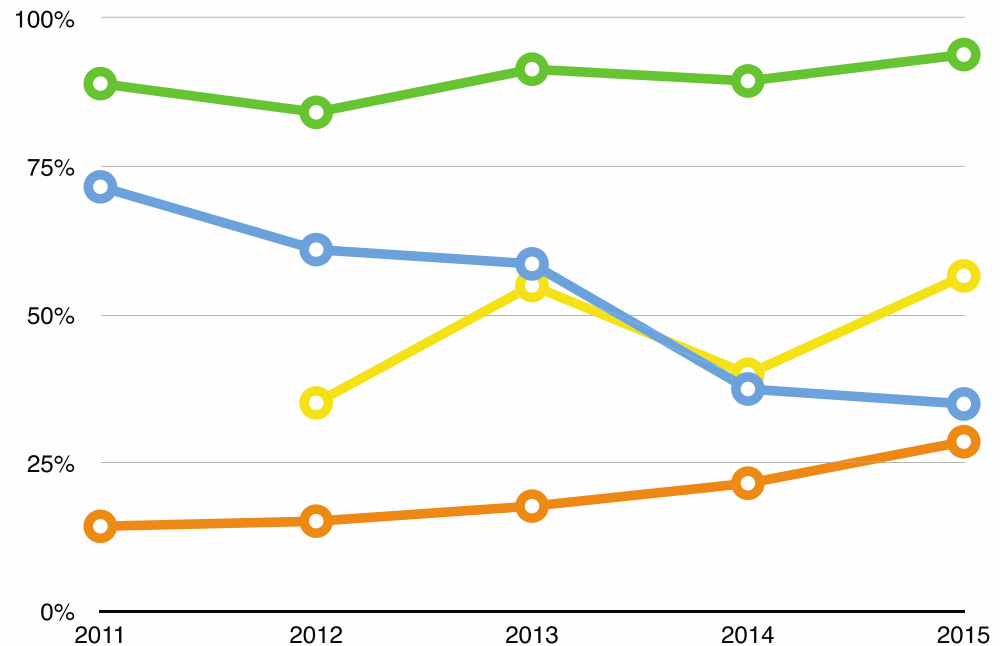


- Given the limited size of the disk buffers in O2 and Tier 1s, all CTF data collected in the previous year, will have to be removed before new data taking period starts.

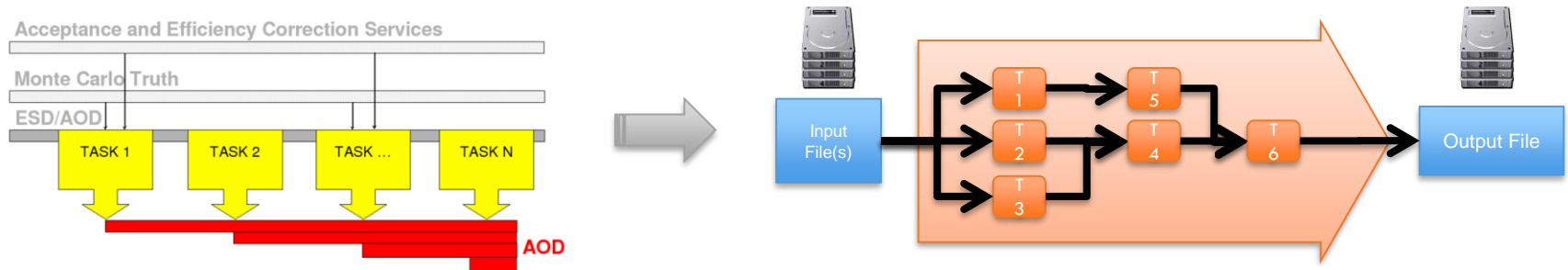
Optimizing Workflow Efficiencies



Wall time / CPU time











New in Run 3: Analysis Facilities



- Motivation
 - Analysis remains /O bound in spite of attempts to make it more efficient by using the train approach
- Solution
 - Collect AODs on a dedicated sites that are optimized for fast processing of a large local datasets
 - Run organized analysis on local data like we do today on the Grid
 - Requires 20-30'000 cores and 5-10 PB of disk on very performant file system
 - Such sites can be elected between the existing T1s (or even T2s) but ideally this would be a purpose build facility optimized for such workflow

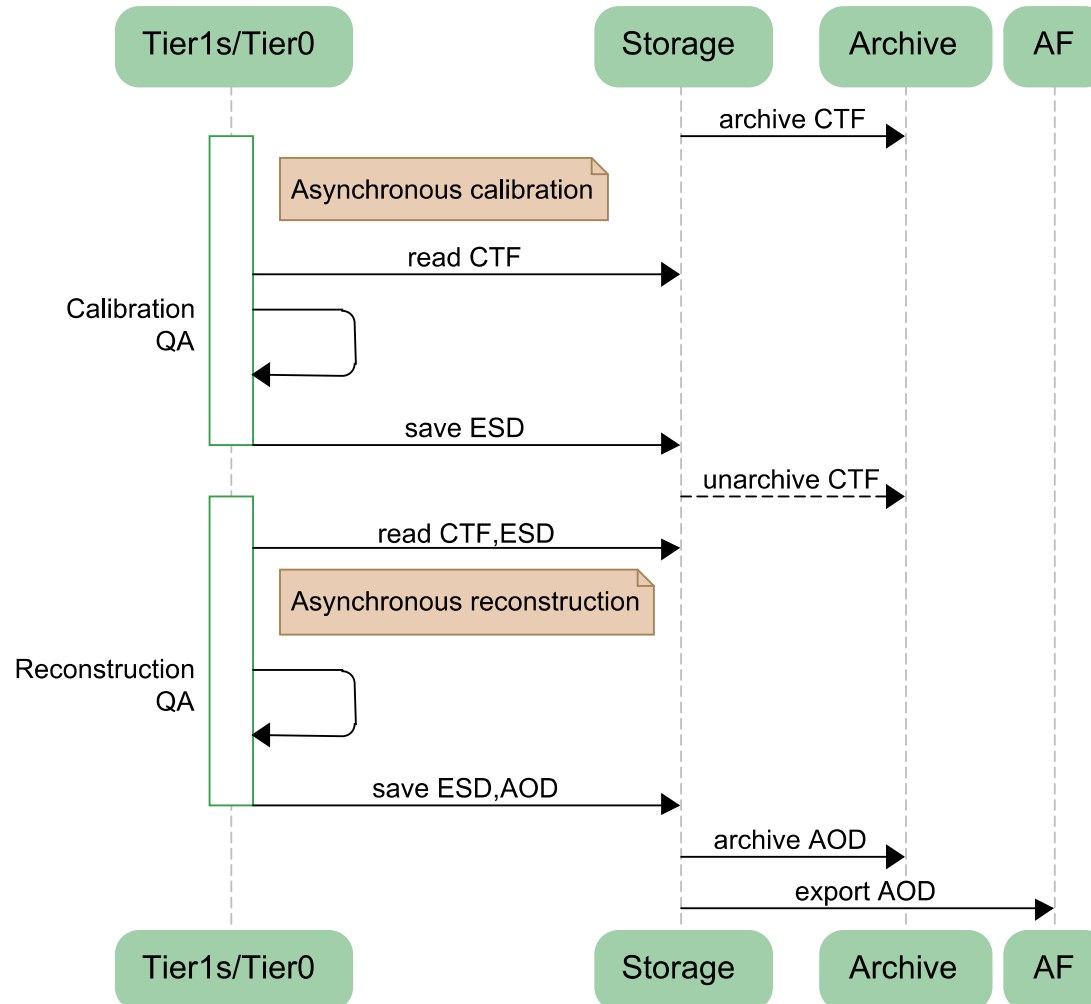
Agenda for today

Convener: Predrag Buncic (CERN)		
09:00	Analysis in Run 3: Constraints and Improvements Speaker: Predrag Buncic (CERN)	20m 
09:20	Current practices: Using non standard "nano" AOD for J/Psi analysis Speaker: Ionut Cristian Arsene (University of Oslo (NO))	20m 
09:40	Current practices: Using ESDs for V0 analysis Speakers: Roman Lietava (University of Birmingham (GB)) , Roman Lietava (Birmingham)	20m 
10:00	Current practices: Heavy flavor AODs Speakers: Andrea Festanti (CERN) , Andrea Festanti (Universita e INFN (IT))	20m 
10:20	Discussion	10m 
10:30	Coffee	20m
10:50	Nano AOD and how to use them in Analysis Trains Speaker: Markus Bernhard Zimmermann (CERN)	20m 
11:10	Data model for Run 3 AODs Speaker: Mikolaj Krzewicki (Johann-Wolfgang-Goethe Univ. (DE))	20m 
11:30	Discussion	15m 



Backup

Tier 0/1 Workflow



Tier 2 Workflow

