Data flow and data formats
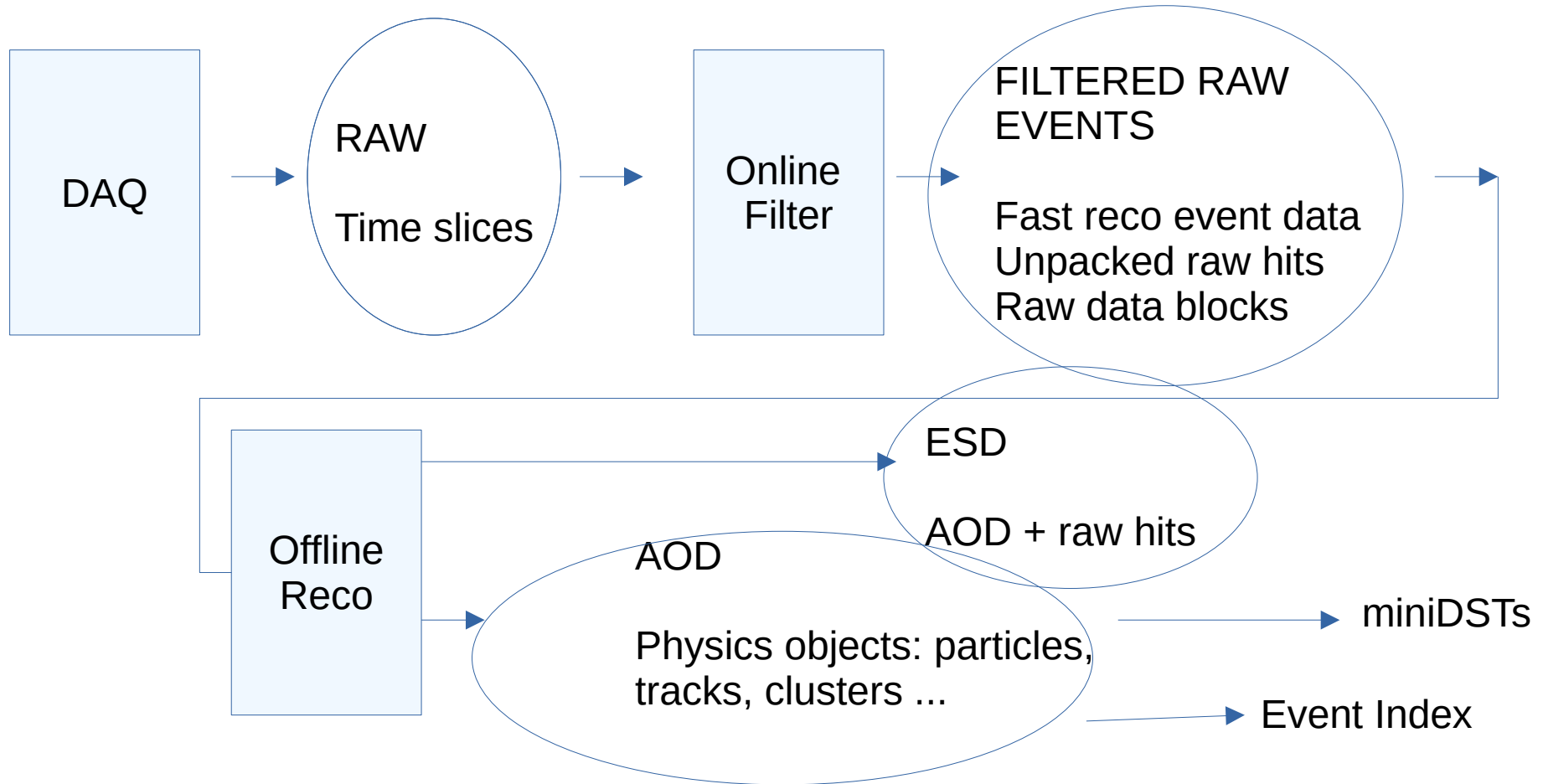
Alexey Zhemchugov

08.02.2022
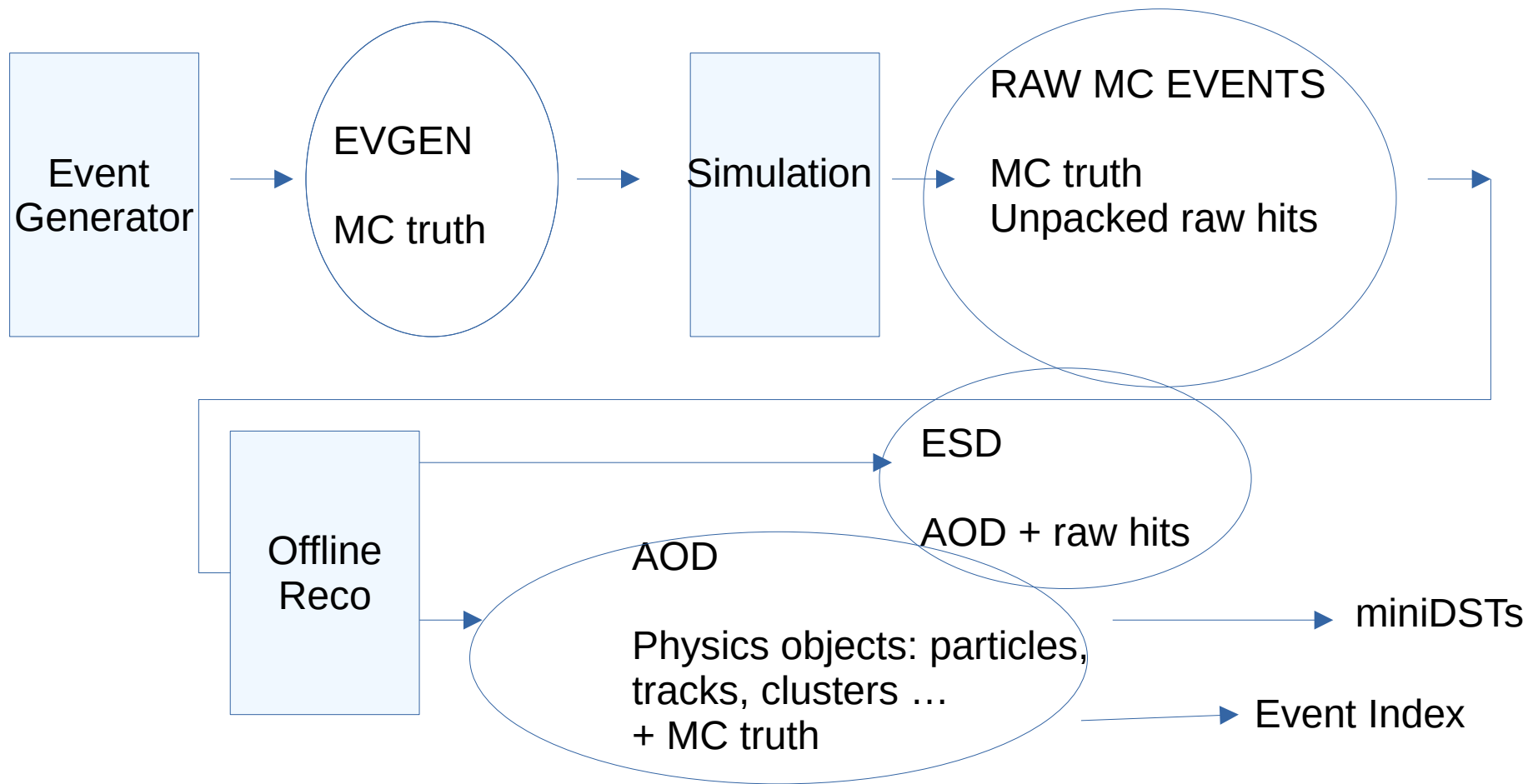
Data flow

- Real data
- MC data
- ML training
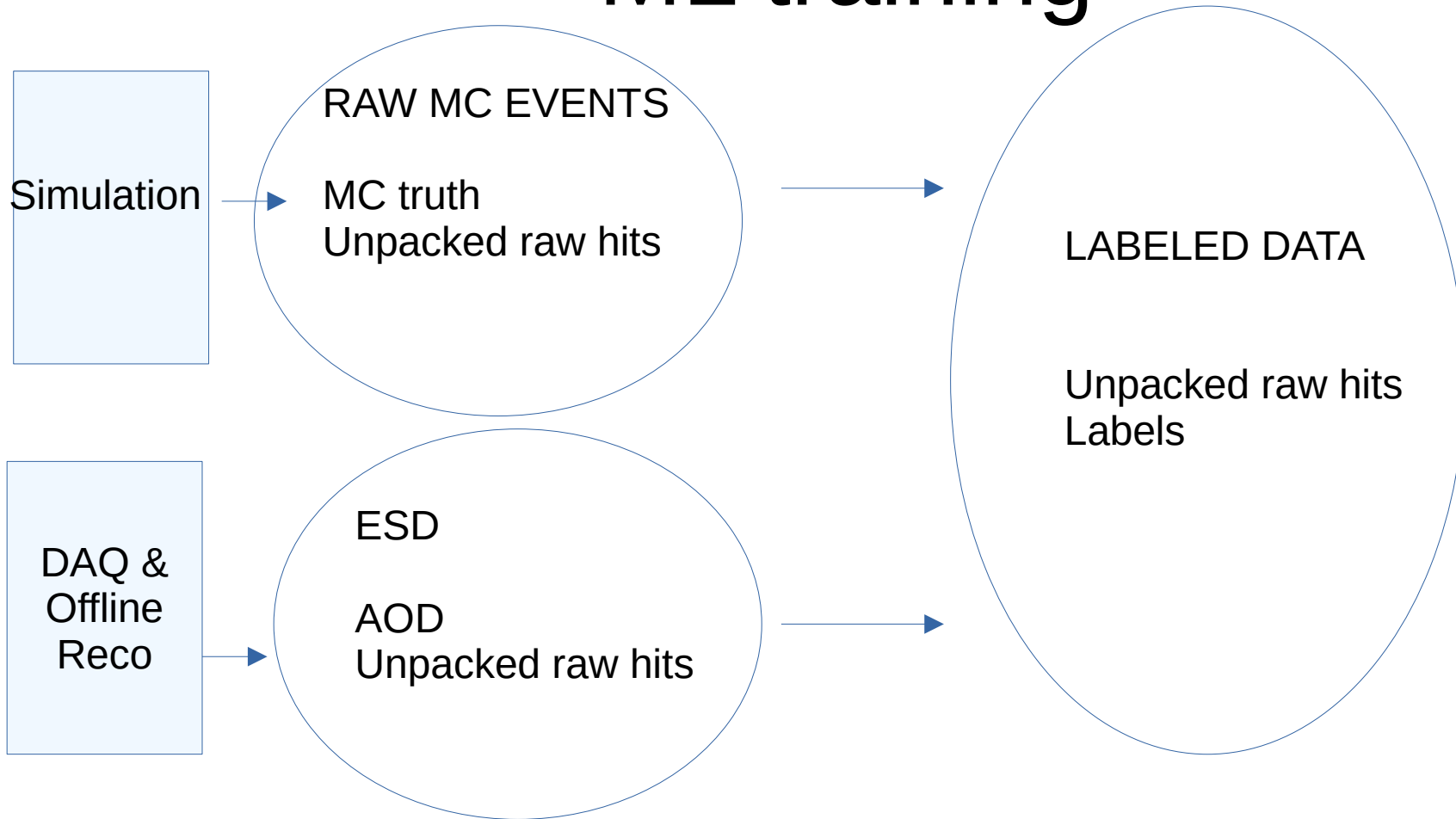
Real data

DAQ → RAW / Time slices → Online Filter → FILTERED RAW EVENTS

Fast reco event data
Unpacked raw hits
Raw data blocks

Offline Reco → ESD / AOD + raw hits

AOD

Physics objects: particles, tracks, clusters ... → miniDSTs

→ Event Index

MC data

ML training

Simulation

RAW MC EVENTS

MC truth
Unpacked raw hits

DAQ &
Offline
Reco

ESD

AOD
Unpacked raw hits

LABELED DATA

Unpacked raw hits
Labels

Data formats

- *RAW (determined by the DAQ*

- FILTERED RAW EVENTS

- AOD

- ESD

- EVGEN

- RAW MC EVENTS

- LABELED DATA

We need the definition of each data format and transformations to process one data format to another!

You are welcome to join this task force

AOD structure (draft 0)

- **Event header**
  - Run ID, FrameID, BCID, Event ID (?)
- **List of Vertices**
  - Vtx,y,z + list of particles
- **Particles**
  - ParticleID, list of tracks, list of ECAL, RS, ZDC clusters, TOF, AEG, BBC hits + MC truth for MC
- **Tracks**
  - (5 track parameters, error matrix) x (start of track, end of track) x (3 particle types), dEdx
- **ECAL clusters**
  - X,Y,Z, energy, error, cluster shape (?)
- **ZDC clusters**
  - X,Y,Z, energy, error

- **RS clusters**
  - X,Y,Z, energy, cluster shape (?), prob_muon ?
- **TOF hits**
  - X,Y,Z, TOF
- **AEG hits**
  - X,Y,Z, Amplitude

Should we keep it in AOD or just have only established PID instead?

- **BBC hits**
  - X,Y,Z, Amplitude
- **Polarization**
  - Polarization degree ?

Every extra byte will become one extra Terabyte per year!