Introduction
High resolution detectors in high energy nuclear physics deliver a huge amount of data which is often a challenge for the data acquisition and mass storage.
Lossless compression techniques on the level of the raw data can provide compression ratios up to a factor of 2. In ALICE [1], a data compression technique has been developed for the Time Projection Chamber (TPC) [2] to reach an overall compression factor suited for data taking in Heavy Ion collisions. Studies on simulated TPC data have earlier shown the potential for data compression [3, 4]. The application on real data is presented for the first time. With a combination of on-line cluster reconstruction from raw data and a subsequent lossless data compression, the data volume can be reduced by a factor 4 to 6 depending on the data sample.

Data compression scheme
Approach: Replacing detector raw data by pre-processed data from the High Level Trigger (HLT)

Online systems

- Detector
- HLT

Offline

- Storage
- Reconstruction

⇒ HLT data can enter offline detector reconstruction at different stages

Model for TPC raw data - Clusterization
Time Projection Chamber provides simultaneous measurement of three-dimensional track information:
- 3D readout of segmented pad planes
- Sampled drift time of induced charges
- Measurements from clusters, which are calculated from the raw data as first step in the reconstruction.
- Clusters distributes over several readout pads.HLM required precision for tracking

⇒ single point clusters: determined pad residuals, not used in tracking but useful charge information for PID

In the HLT the cluster reconstruction is performed in an FPGA co-processor in real time [5, 6].

Data reduction in clusterization
The reconstruction of clusters from raw data is a lossy transformation, the original raw data can not be restored. However, the loss in the required information for the event reconstruction is negligible for the physics analysis. A small data reduction factor is achieved in this step.

⇒ data reduction by factor 1.1 to 1.5 depending on the event size

Raw cluster format
The reconstructed clusters are stored in raw coordinates which makes the recorded compressed data independent of drift time calibration and further corrections.

Conclusions
- Effective data compression is achieved by a combination of lossy transformation (cluster reconstruction) and lossless data compression
- Data compression solution for TPC data operational in ALICE since 2011 Pb-Pb data taking.
- 2011 data dominated by large events, average compression factor of 4.4 has been measured.

References