

# **Data Processing for CLAS12**

**Gagik Gavalian (Jefferson National Laboratory)**

# WHY ?

## HIPO:

- Custom Data format was introduced to deal with our daily needs.
- Compression is needed for large scale data (RG-A spring has 100 time more data than G11)
- Debugging requires fast random read data format, multithreaded enabled.
- Bucket/Record Tagging ability for big data.
- Was designed having our needs in mind.

## COULD WE USE A STANDARD FORMAT ?:

- EVIO - NO random read for files >2GB, NO compression.
- LCIO - been told by people who use it - "Don't use it"
- APACHE - slow readout time, NO random read
- HDDM (GlueX) - NO random read, NO multithreading, NO JAVA interface
- HDF5 - NO native JAVA interface, NO predefined structures, SLOW compression algorithm

# News

## HIPO-4 TRANSITION:

- Forth generation of HIPO library is introduced to CLAS12 reconstruction environment.
- File index structure is modified to include tag information for each record. Significantly improved IOPS.
- Writer allows multi-tag output, flexible configuration for basket sizes for each tag.

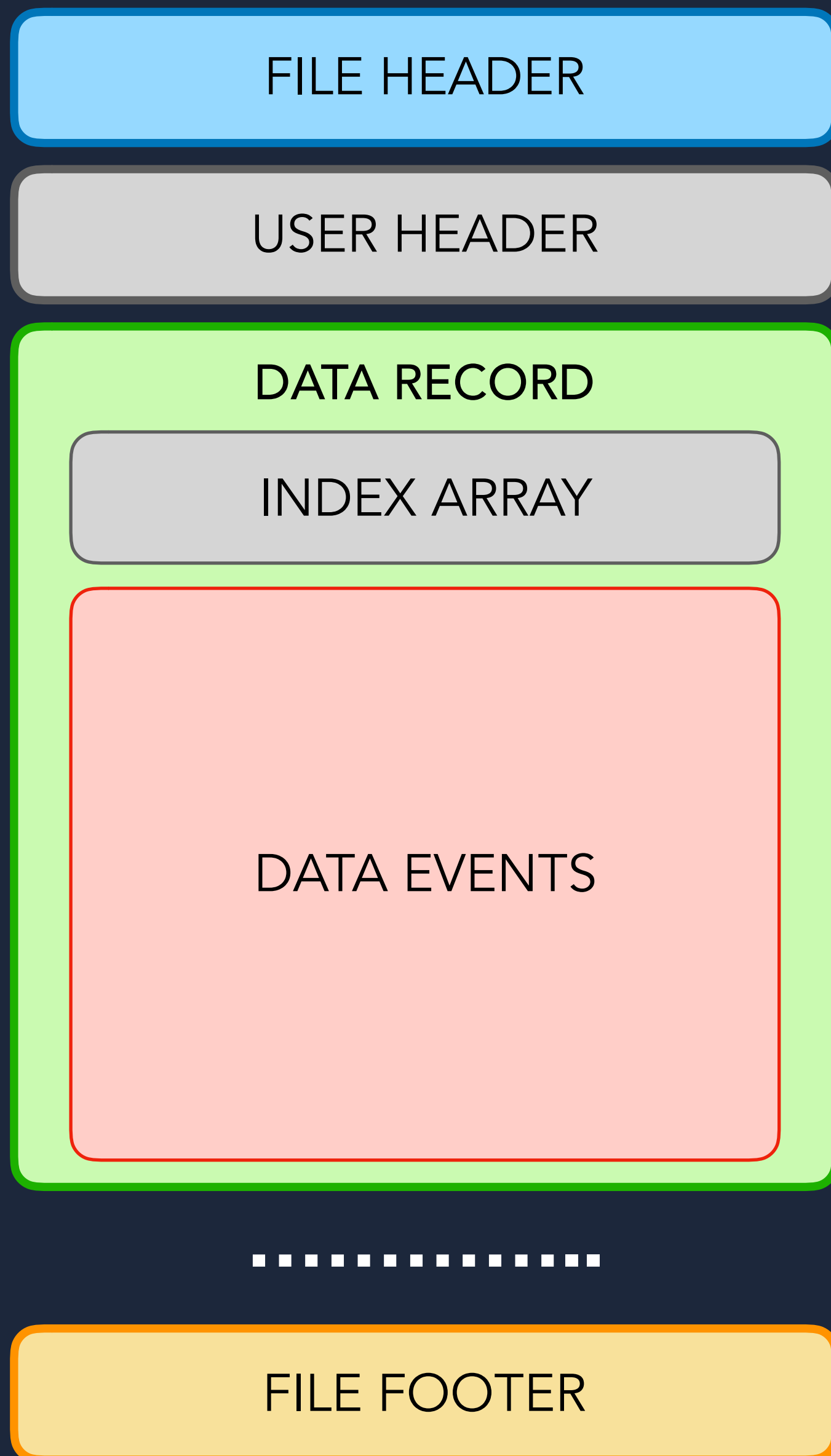
## PROGRESS:

- Decoder conversion is complete.
- Reconstruction engine conversion is complete.
- CLARA/IO services updated to HIPO-4
- DataSource interfaces are switched to HIPO-4
- New DataSource interfaces are implemented for backward compatibility with HIPO-3.
- C++ Reader for HIPO-4 is complete.

## TO DO:

- Convert CLARA Trains IO services to HIPO-4.
- Implement FORTRAN wrapper.

# HIPO Data Format (File Format)



## User Header

Contains information about the record dictionary, format. User specified parameters related to conditions of the experiment.

## Data Record

Compressed buffer of data consisting of events and index. Record header provides number of events and the TAG for the record. Data records are typically ~8 MB.

## Index Array

Array of event offsets inside the event buffer. Dynamically creates event random access table.

## FILE FOOTER

Contains positions of every record in the file with number of events for fast random access. Also has tags for each Data Record.

# Data Formats

## DATA RECORD TAGGING

- Tagging mechanism for events inside the file.
- SCALER/EPICS events are assigned unique tags, they appear in separate records and can be read separately from other events.
- Events topologies will also be tagged for considerable speed up in skimming stage.

## UTILITIES

- New file filtering mechanism is implemented with regular expressions.
- Filtering files with expression parsing is being developed.
- Merging files now preserves the tagged structure of input files, other tags can be also introduced.

### EVENT TOPOLOGY FROM CLAS12 (RUN #3856)



**59.7%** Trigger particle is not an electron.  
No electron Forward Tagger.

**25.6%** Electron trigger.  
Forward Detector

**14.7%** Forward Tagger  
No Electron in ECAL

# BACKUP SLIDES