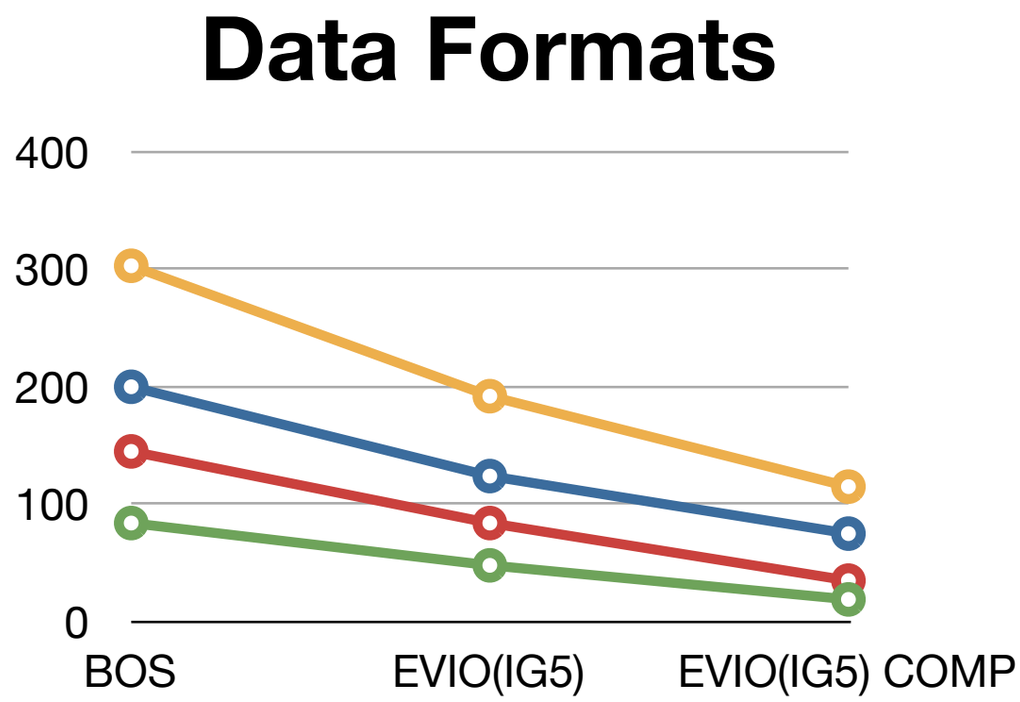
Data Storage and Compression

- **Data Format:**
  - new data format is developed for row wise bank storage (iG5).
  - interface exists for both Java and C++.
  - dictionary describes bank structure and data types.
  - the raw data with it's dictionary is stored in EVIO buffers.
  - lossless compression reduces the raw data size by 40%-60%.
- **CLAS6 Data storage (Data Mining):**
  - experimental data from clas6 converted into iG5.
  - different experiments have different bank structures.
- **Compressions:**
  - data converted to iG5 from BOS shows reduction in file size.
  - further compression reduces the data size by factor of 3.



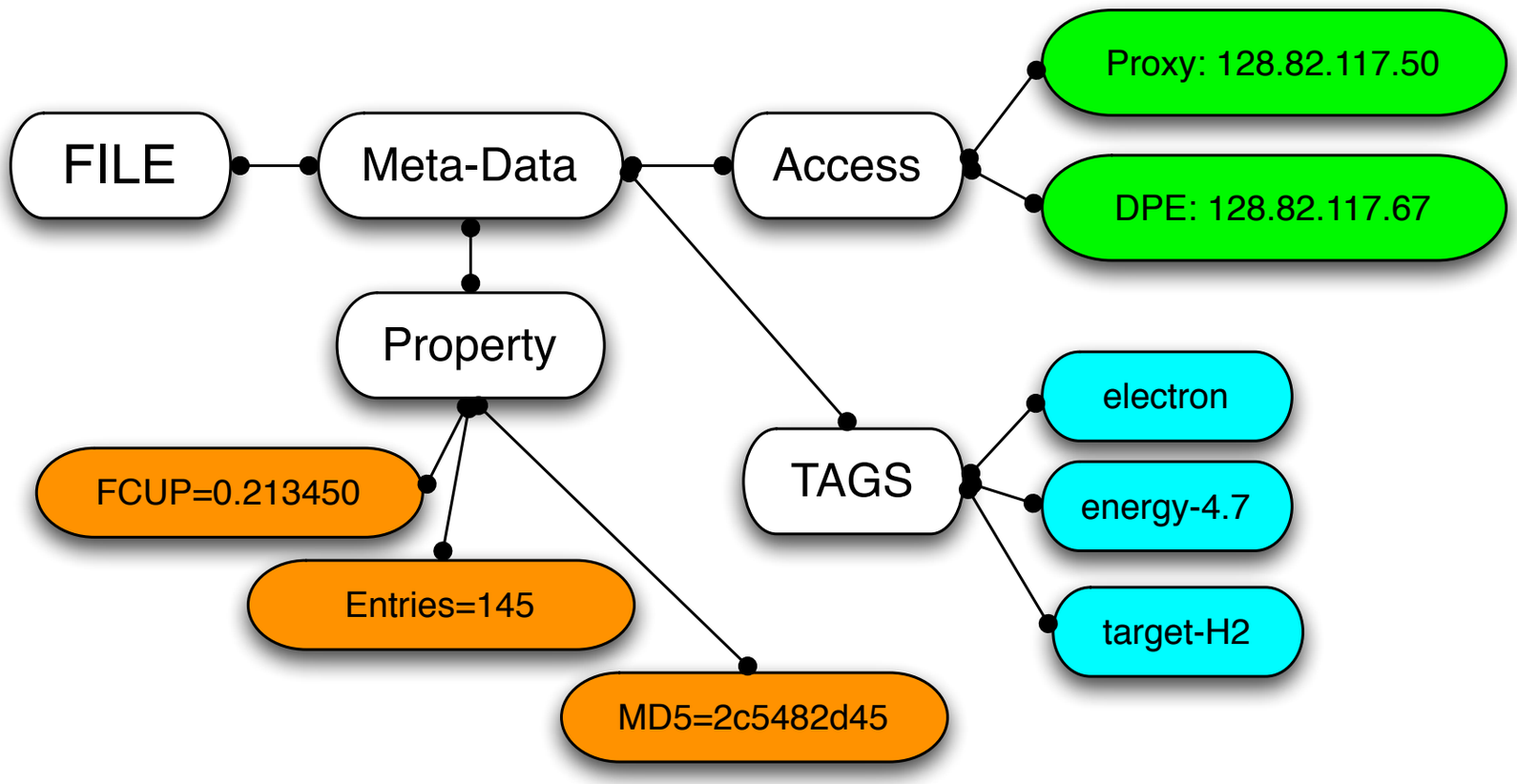
Data Format

- **Data Format Collection:**
  - data is stored in buckets.
  - each bucket contains ~5000 events.
  - buckets are compressed using GZIP.
  - compressed bucket is stored as a EVIO leaf.
- **Data File Header:**
  - header contains dictionary for the banks.
  - header contains EPICS information (FCUP,TIME,FLUX)
- **Streaming File Content:**
  - data streamed to the service contains iG5 ByteBuffer.
  - description string contains the dictionary.
  - attributes contain EPICS information.

Data Distribution and Discovery

- **Tagged File System (TFS):**
  - provides meta-data for each file on the server.
  - files are discovered using associated tags, related to experiment conditions.
  - properties provide related information about experiment.
  - access points provide proxy locations for data (multiple-sources).

- **Meta-Data descriptors:**
  - meta-data descriptors are downloaded from service providers.
  - a synchronized list is created on the client side.
  - data discovery crawlers will be implemented to keep track of available data.



Tag-FS Associated Tags

- With each choice of specific tag, all associated tags are presented to the user.

Selected Tag **Associated Tag**

eg2 | target-Fe56 | target-D2 | beam-electron | 4.7 Gev

eg2 | target-c12 | target-D2 | beam-electron | 5.7 Gev

beam-electron | 4.7 Gev | e2b | target-c12

target-H2 | beam-photon | g11 | 4.0 Gev

Tag-FS Associated Tags

- Choosing additional tags reduces the number of associated tags.
- Groups of files can have chosen tag combination.

Selected Tag **Associated Tag**

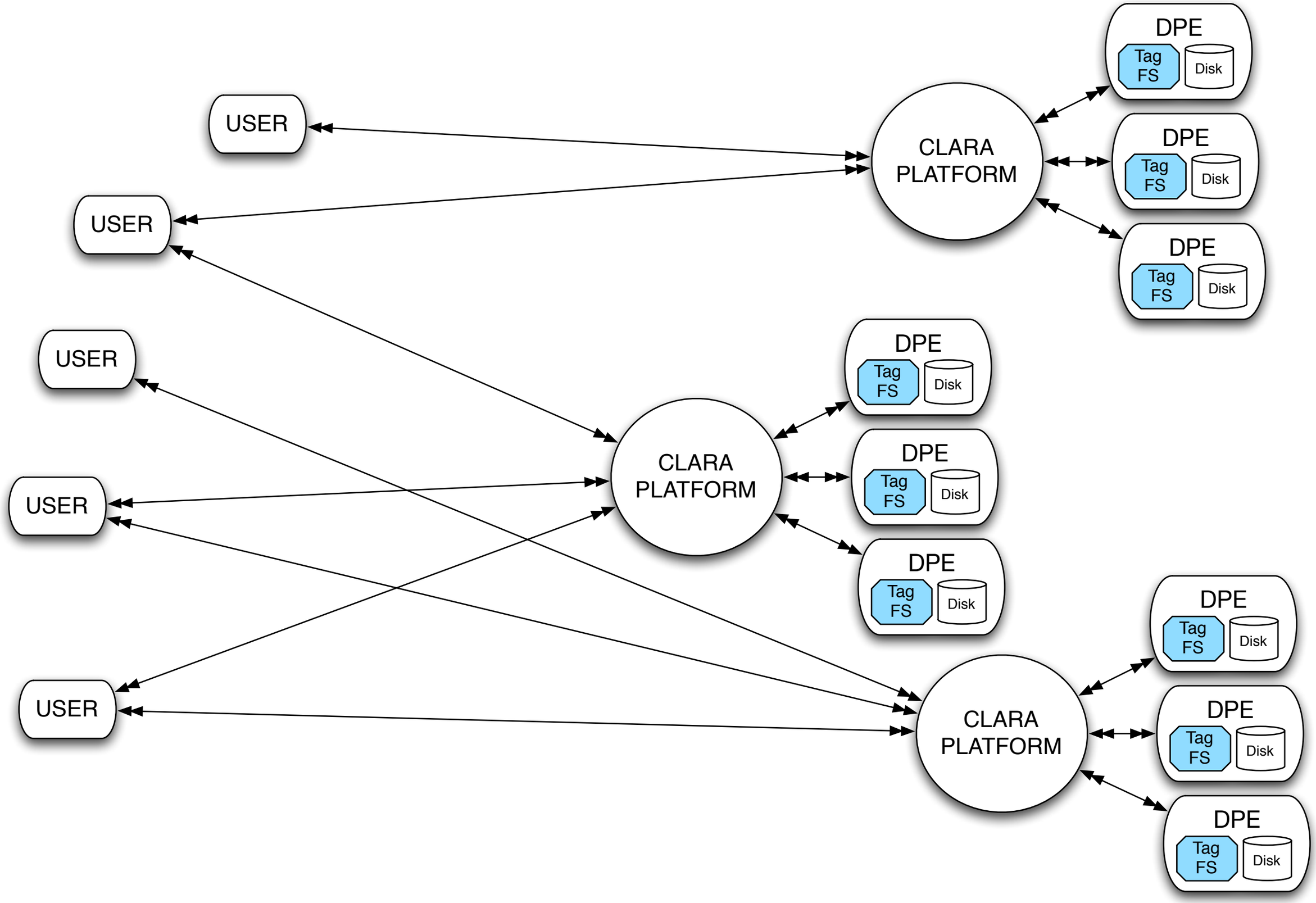
eg2 | target-Fe56 | target-D2 | beam-electron | 4.7 Gev

eg2 | target-c12 | target-D2 | beam-electron | 5.7 Gev

beam-electron | 4.7 Gev | e2b | target-c12

target-H2 | beam-photon | g11 | 4.0 Gev

Tag-FS Distributed System



Ideas

- **Separate files.**
  - **different topologies can be written in separate files.**
  - **each file can have a tag with topology description.**
  - **tagfs search will allow analyzing similar topologies from different runs.**

- **Rearranging File:**
  - **different topologies can be grouped together.**
  - **file header can contain a map of topologies with event ranges.**