DØ Note 3466
CD Note TN0089

Glossary of terms for the Sequential Access Model Data Access System

June 15, 1998

Abstract

In the Sequential Access Model (SAM) Data Access Working group we are developing Requirements and Architectures for a Data Access System. This necessarily involves terminology. To facilitate communication and clarify our own use of terms we have developed a Glossary of commonly used terms.

1 Glossary

The goal of this glossary is to define terms used in the Data Handling Projects for Run II, to provide a reference for checking consistency of use of these terms in the documents, and to provide a more precise description of these terms within the their context. There is an attempt to define these terms in relation to the physics analysis domain in which they are being used.

- **Analysis Group** - Group of physicists who organize to coordinate their access the same dataset or files for the purpose of analysis.
- **API** - Application Programming Interface. Set of callable routines through which access to event data can be controlled and data read or written.
- **Application Framework** - Software to whose API a user or physicists application code code is written, and under which the program is executed.
- **ATL/Robot** - Automated Tape Library where the mounting and dismounting of tapes in the tape drives is done without human intervention though an API.
- **Auditing** - Information gathered about the processes running and the parameters associated with them, the transactions executed, and the user requests queued and serviced.
- **Bookkeeping** - Stored information describing the production, location and history of files.
- **Buffer** - Temporary disk storage of data to smooth the flow of data when the instantaneous rate of the input and the output are different. The policy of what data is written to a buffer does not include the assumption that it will be reused - cf. Cache.
- **Cache** - disk storage used to store more frequently accessed data in order to reduce the latency and increase the throughput of access. The Cache reduces the effective access time to data stored on tape. Cached data, by definition, is expected to be a subset of the data that will be (frequently) reused and whose size allows it to be stored on the available disk.
- **Cache Policy** - decisions on how to manage the cache.
- **Catalog** Database of semi-static information about Events or Files that does not include the actual Event or File within it. A catalog does not necesarily include the location of the event or file.
  - **Event Catalog**
  - **File Catalog**
- **Consumer** - User Process that receives data from cache or tape.
- **DA** - Data Acquisition, collects the raw data from the detector
- **Data** - In the context of SAM data and event are used interchangeably. In general if the term Data is used it could additionally apply to calibration constants, meta-data that applies to more than a single event, detector or run conditions that apply to a whole run or sequence of runs etc.
- **Data Streams - Event Streams -** Organization of Events based on anticipated access patterns, such that access to the data is facilitated and can be achieved as fast and as efficiently as possible.

  - **Physical Data Streams -** Data Streams determined by the Trigger and Level3 Event information and physically located together on disk or tape.
  - **Logical Data Streams -** Data Streams determined by the expected selections and queries used to access the data (e.g 4 jet stream).

- **Database** - Records of information organized for rapid search and retrieval.

- **Database Management System** - (DBMS) A set of programs that control the organization, storage and retrieval of a database.

- **Dataset -** Collection of events that has a meaning from a Physics perspective. A dataset can be of varying types and sizes - e.g. Raw Dataset - collection of Raw Events; Top Dataset - collection of events that are Top Candidates; Run Dataset - collection of events from a single data acquisition Run; Run Ia Dataset - all events from Collider Run Ia; Calibration Dataset - related collection of calibration events.
  - **Primary Dataset -** A dataset output from the initial production or farm processing.
  - **Derived Dataset -** A dataset obtained from further processing of events after initial reconstruction on production systems.

- **Data Tier -** Event data is classified according to its content into tiers where each tier holds a particular subset of the whole event - e.g. Raw Event, Analysed Event. Lower tiers contain more information for each event - e.g. the RAW event data and higher tiers contain a refined level of event information - e.g. EDU50. The top most tier contains an event catalog.

- **Event -** Any information associated with a single Collision in the Experiment Detector for which data is read out and archived.
  - **Raw Event -** Event information as recorded from the online/level 3 system.
  - **Reconstructed Event -** Event information after processing by the reconstruction program and production systems.
  - **Analysed Event -** Event information output from one or more physics analyses being performed on the event.

- **Event Clustering -** see Data Clustering

- **Event Data Unit -** Event Data Unit (EDU) is data identified by a unique combination of Run Number, Event Number, Type of Event Data Unit. Any type of data record or data from any data tier, which can be read as a unit from some physical devices and is associated with a particular event.
  - **EDU50 -** Event Data Unit of size around 50KBytes.
  - **EDU250 -** Event Data Unit of size around 250KBytes. This is expected to be the size of a Raw Event.
  - **EDU5 -** Event data unit of size arund 5KBytes. This term is useful as it is expected that enough spinning disk will be available to store all Events if the size per event is around 5KBytes. It is expected that Thumbnails will be EDU5s.

- **Event Management System -** Hardware and Software System to manage and coordinate access to the meta-data and events.

- **Event Picking or Pick Event -** "Random access" user requests for particular events or selections of events. This mode of event/data access can result in vastly different patterns of access to the data than the Freight Train and Production data access modes. Not understanding the quantity of and need for this type of event access could result in constructing a system unable to deliver the needed throughput in this area.

- **Export** - Making tapes previously available through the ATL or Mass Store no longer available, and only available external to the system.

- **File** - Collection of bytes on disk or tape treated as a single unit and identified by a file name and physical location. A file may be a program, a document, a database, or some other collection of bytes.

- **File Family** - Collection of files distributed across more than one tape volumes and potentially written in parallel to more than one tape volume. Once a volume contains data from a file family, the rest of the volume will contain only files from that same family.

- **Freight Train -** Mechanism and process by which large datasets are processed in a coordinated fashion. A Freight Train consists of the organized delivery of a large dataset over a period of up to several weeks or months to multiple analysis programs. The data provided by a Freight Train is delivered at a rate determined by the data access and delivery system and not by the ability of the analysis programs to absorb and process it. Thus, a more cpu intensive analysis program may not receive all the data of the dataset in a single passage of the freight train. Mechanisms are in place to monitor which events/files are successfully accessed, and for resending those that were missed.

- **GUI -** In the context of SAM, the Graphical User Interface is imagined to provide integrated control and monitoring of the user requests and the processing system that will satisfy them.

- **HSM -** Heirachical Storage Management System. Provided by that Run II Joint Project. SAM is a user of the Run II HSM and does not construct or operate it.

- **Import** - Making tapes available through the ATL or Mass Store that were previously outside access through and control by the system.

- **Latency -** Time between request for data and the delivery of data to the requesting program.

- **Luminosity** - a measure of the number of potential proton/anti proton collisions per second. Used to normalize measured event rates to the total proton/anti-proton cross section or to each other.

- **Luminosity Block -** Time stamped record which records the luminosity for a time interval for each trigger.

- **Luminosity Record** - data which summarizes the luminosity integrated over a given time period. Different triggers may have different luminosities due to prescales and special beam conditions.

- **Map -** Relationship between information in an Event or File Catalog to the Physical Location or other attribute of an Event or File.

- **Meta-data -** Set of information used to describe Events, Files or Datasets for the purposes of allowing location of and access to the data, selection of subsets of the data etc.
- **Monitoring** - Gathering and presentation of information about the running system, the processes executing and the data flowing.
- **MSS** - Mass Storage System - system that manages a large amount (Terabytes or Petabytes) of data on tape and provides for writing and reading of this data. IN general an MSS includes an ATL, can include a hierachy of storage on Roboticly accessibly tapes, tapes residing on shelves and/or caching of the data on local or distributed disk. HPSS - A Mass Storage System in production use at Fermilab. Described at http://www.sdsc.edu/hpss/. Enstore - a Prototype mass storage system at Fermilab, based on the DESY OSM model. Will be described at http://hppc.fnal.gov/enstore.
- **On-Demand Data -** Data access mechanism for moderate-sized datasets to be processed by single users or small groups of users. Usually confined to a small set of files and used by analysis jobs that run in a few hours or days (depending on the total amount of data in the request). One important usage of this access method is the debugging of code for analysis jobs, such as for eventual use with Freight Trains.
- **Physics Data Stream - Physics Event Stream -** Collection of data, specified by the Trigger value. In D0 there will be up to 15 Physics Data streams.
- **Pick Event - see Event Picking**
- **Post-reconstruction analysis and production -** Processing done on a Primary or Derived Dataset.
- **Processing Unit -** Program that performs calculations on a Dataset.
- **Production Farm -** System of physically independent computing resources used in a coordinated fashion through the use of control software, for the reconstruction or analysis of event data. Typically Production Farms have been used for the processing of the Raw Events to produce output Data Summary Events, and are best suited for situations in which the "CPU used per event" is much larger than the I/O needs per event. Nowadays, production farms are used for reprocessing of data or "production analysis" where a large body of any Event Data must be processed by a single executable.
- **Project** - Organization of coordinated processing of files this will occur in a coordinated fashion such that the events are read only once from the storage management system and are delivered in parallel to all the analyses programs that want to access them.
- **Query Optimizer -** A program and/or business process that receives user requests and attempts to evaluate how long they will take and which h/w resources the request will require. It is planned to merge the requests from many users to optimize use of the resources.
- **Reconstruction** - Processing of the raw data, and associated calibration data, to make one or more derived data sets.
- **Request Manager -** (is this true?) Program that handles the user requests and allocates a "token" to authorize a program to access and obtain resources or data.

- **RIP** - Reconstruction Input Pipeline project which provides data from the data acquisition system buffers to the Mass Storage System, and from the Mass Storage System to the Production Farms.

- **Run** - Run I, Ia, Ib, IIa, IIb, etc identify periods of Collider operation - typically of months to a years duration. A data acquisition Run identifies a set of data acquired sequentially with a single set of calibration and detector constants. A Run is identified by an increasing Run Number which is typically stored in each Raw Event. The Run Number is used to identify temporally connected EventData. In stable detector and collider operation, Runs last between a few hours and a day.

  - **Physics Run** - Sequential set of events identified by a single number known as the "Run Number". In general, all events from a single physics run are associated with a single set of calibration data.

- **Run Conditions** - Parameters associated with a Physics Run e.g. Luminosity, state of the detector elements etc.

- **Shelf (Storage) -** Tapes not currently stored in the Automatic Tape Library or Robot. It is expected that the HSM system will provide mechanisms for moving tapes in both directions between Shelves and the Robot(s).

- **Station** - Computer system on which data processing and/or analysis is executed.

- **Storage Management Software (SMS)** Software to manage the Heirachical Storage Management System (HSM).

- **Thumbnail -** A Thumbnail (Event) is sufficient on which to perform most physics analyses. It is expected/hoped that the experiment can define the size of the Thumbnail such that the whole of the Run II dataset can be stored on attached disk.

- **Token -** A program requests and obtains a "token" in order to gain authorization to use a resource and access/retrieve data. Once the use of the resource is complete the program must return the token so that it may be allocated to another user/program.

- **Warehouse -** Sum of all the stored data.