Table of Contents

Some ideas for writing the Physics Analysis Work Book

Introduction

Here are some thoughts about the PhysicsAnalysisWorkBook

- We should structure it in a way that mirrors a person's thought process when they start to think about doing analysis
- We should provide a lot of examples, both simple and complex, with annotated python scripts, log files, source files, etc.
    - AnalysisSkeleton.cxx is a good file to orient people
    - Other examples can include topics like, how do I vertex a selected set of tracks, etc
    - There could be links to other standard software, e.g., RooFit⬀ (a layer on top of ROOT to do things like multi-variate likelihood fits), neural nets, etc.
- We should have a list of people who are willing to answer questions in a timely manner, e.g., analysis tools experts, physics group conveners, etc. In addition, also provide a FAQ
- ...

How do I run on data (or MC)

- What formats are available? Point out the pros/cons of each format.
    - For instance, one can run on AOD's: Pro – most pristine format with no cuts on objects, Con - this will require working in the Athena framework.
    - Root-tuples (whether AthenaAware or not): Pro – no knowledge of Athena framework required, Con –were cuts made to produce them?
    - Is there a preferred format? Point to tools like EventView⬀ or whatever other such tools might exist
    - ...

- What information is stored in these formats?
    - What objects are stored? What variables for each of the objects? An explanation will be very useful.

- Examples on how to run on the formats described above. Here they can just run interactively on a handful of events and look at the output.

- How do I look at some event displays?

- Here we can have more sophisticated tools, e.g., EventView⬀ or whatever comes next. We should encourage authors of these tools to have good documentation, e.g., if they introduce new variables, then there should be a good explanation of them

- How do I include trigger information in my analysis? What triggers are available?

- How do I run on a large(r) sample of data (or MC)?
    - Where do I run batch jobs? What do I need to do before I can do this, e.g., do I need a grid certificate, accounts on the "nearest" Tier 1 and/or Tier 2 centers?
    - Provide sample batch scripts (see Jesse Ernst's script to split jobs and submit them via condor at BNL)

♦ Alternatively, one can submit jobs to run on the Grid. Has this been tested? Does it work?

Data Samples

- Where do I find data samples? We should distinguish between small test samples and large scale running.
  - ♦ Explain how to find test samples. Should people use dq2_ls?
  - ♦ I tried using the dq2 dataset browser and I found many datasets were empty. In cases where these datasets did contain files, I got some cryptic error messages – this needs better documentation
    - ◊ Someone remarked to me, "The Panda Monitor" has a high degree of color coding, but sometimes it is not clear what things mean.
- What about running on large data samples?
  - ♦ Is it better that people first get in touch with physics conveners or should they go off on their own?

MC samples

- How do I generate small test samples
  - ♦ Very useful to produce Evgen files to ensure that I am setting up things correctly. How do I run on these files?
  - ♦ Some people may need to filter events when generating events. It seems that there is no centrally approved software to do this and that everyone is writing their own filters?
    - ◊ Is that what is envisioned?
    - ◊ Can MC production farms deal with these filters?
  - ♦ Produce AOD's (or the preferred format) – already explained how to do this in the standard Work Book or above

- What about large samples? How do I go about getting a large sample of events? Are there already there? Perhaps it is best to go through physics conveners.

This topic: Main > PhysAnaWorkBook
Topic revision: r3 - 2006-11-29 - VivekJain