

# Streaming Evaluation Tools

Adam Lyon and Jeremy Simmons  
For the Analysis Tools Group  
Oklahoma D0 Workshop Streaming Session  
7/11/02

# How to evaluate streaming

---

## ◆ Assumptions:

- ❖ Analyses are trigger based
- ❖ Commissioning triggers and special runs will not intermingle with physics streams

## ◆ Evaluation:

- ❖ Physical stream sizes
- ❖ For a given set of triggers (representing signal or background):
  - Fraction of the total data that needs to be processed
  - Fraction of the data in stream(s) from those triggers
  - Fraction of tapes that need to be mounted

# Difficulties

---

- ◆ Current data do not represent “normal running”
- ◆ Current trigger lists do not represent “normal running”
- ◆ Tape usage would need to be simulated (try out file families)
  
- ◆ Two solutions:
  - ❖ Things are the way they are – evaluate streaming within current conditions with data and trigger lists we have now
  
  - ❖ Simulate future conditions (e.g. use offline objects to guess at trigger decisions – since we care about primal objects, this wouldn't be too inaccurate). But we still don't have all data objects correct (e.g. missing central muons). Use some Fast MC?

# Implementation

---

- ◆ Run over events to determine trigger-primal streams and trigger set frequencies [Jeremy]
  - ❖ Based on Michael Begel's code
  - ❖ Gets information from the SAM database (no data files needed)
  - ❖ Can process single runs, multiple runs (within same trigger list) and all runs in a trigger list
  - ❖ Queries database for trigger → primal stream  
     $\mathbb{V}\Delta\text{mua\_CJT3: tau, muon}$
  - ❖ Queries database for trigger set frequencies  
     $[\text{JT\_45TT\_L2, JT\_65TT}]=30,$   
     $[\text{mu2ptxbttx\_fz, mu2ptxwttx\_fz}]=760, \dots$
  - ❖ Results are saved to an XML file

# EventsInfo XML

---

- ◆ Can be run outside of D0mino
- ◆ Can change the trigger → primal stream mapping
- ◆ ATG can store events XML in a central place
- ◆ Perhaps a fast MC can generate events (more “realistic” runs than current)

# Processing

---

- ◆ Determine the make up of the primal streams

muon: [mu2ptxbttx\_fz, mu2ptxwttx\_fz]=760, [mu\_hi]=100, ...  
electron: [ele\_hi\_2, ele\_hi]=30, [ele\_lo, ele\_hi]=400, ...  
muon+electron: [mu\_hi, ele\_hi\_2]=20, ...

- ◆ Determine physical stream destinations

- ❖ Uses the same XML as Jon for the Stream Scheme

```
muon? <
  +MUON
  electron? <
    +ELECTRON-MUON
    jet? <
      +JET-ELECTRON-MUON
      tau? <
        +TAU-JET-ELECTRON-MUON
        -TAU-JET-ELECTRON-MUON
```

# Stream Scheme – more complicated

---

```

                                +TAU+ELECTRON+MUON
tau? <
                                +TAU2-TAU+ELECTRON+MUON
                                tau2? <
                                -TAU2-TAU+ELECTRON+MUON
electron? <
                                +JETLOW-ELECTRON+MUON
                                jetlow? <
                                -JETLOW-ELECTRON+MUON
muon? <
                                +PHOTON+ELECTRON+JET-MUON
                                photon? <
                                -PHOTON+ELECTRON+JET-MUON
                                electron? <
                                -ELECTRON+JET-MUON
jet? <
                                +TAU-JET-MUON
tau? <
                                -TAU-JET-MUON
```

# Processing

---

- ◆ Follow the Stream Scheme Tree and deposit each primal stream set into a physical stream
  - +MUON: muon, muon+electron, muon+jet, muon+electron+jet, ...
  - +ELECTRON-MUON: electron, electron+jet, electron+tau, ...
- ◆ Now have a list of the destinations for each set of events (grouped by trigger set)
- ◆ Scan the physical streams to determine the destination of each trigger – make a table

# Querying

---

- ◆ Given a set of triggers (ands & ors), determine where the events went
  - ❖ How much data in those triggers?
  - ❖ What physical streams did they go to?
  - ❖ What fraction of the data is in those physical streams?
  - ❖ How many events in those streams are interesting (have those triggers)?

```
Physical streams for trigger set: [MEHI_SH] or [MEMN] or [MELO_L2]
+-- Triggers represent 81537 events (20.8% of all 392426 events)
+-- 2 physical streams need to be processed: 228817 events (58.3% of all events)
+-- 81537 events (35.6%) in these streams are interesting
+-- Breakdown of Physical Streams:
+-- For Phys Stream +ELECTRON-MUON: 81290 events of 152376 (53.3%) are interesting
+-- For Phys Stream +MUON: 247 events of 76441 (0.3%) are interesting
```

## What's next

---

- ◆ Interface is very raw
  - ❖ Jeremy and I have ideas for improvements
- ◆ Provide easy way to override primal streams
- ◆ Any other reports needed?
- ◆ Do we worry about file families yet?
- ◆ Have to find more representative runs (if exist)