

Stage 1 (optional)

weekly, hours to complete

Source event data

few TB per billion events

**recompress, select
columns and rows**

Event data (LZ4)

25%-50% of the original

Optional selection and
LZ4-compression step to
reduce overhead.

Stage 2

daily, minutes to complete

batched data

**physics analysis code:
CPU and GPU kernels**

physics results

<1GB

Histograms for statistical
analysis, reduced event
arrays for DNN training.