CLAS12 Computing Resources for PAC Proposals

1 Inputs

$$
\begin{aligned}
T \text{ [kHz]} &= \text{Trigger rate} \\
D \text{ [MB/s]} &= \text{Data rate} \\
P \text{ [days]} &= \text{PAC days} \\
X &= \text{Fraction of triggers that are of interest} \\
S &= \text{Number of simulated events per real event of interest}
\end{aligned}
$$

$T/D/X$ can (and should) be based on existing CLAS12 data when possible. The fraction X includes trigger purity, plus additional criteria applied during analysis.

2 Calculations

$$
\begin{aligned}
V_d \text{ [TB]} &= 0.086 \cdot D \cdot P = \text{Raw data volume} \\
V_r \text{ [TB]} &= 0.4 \cdot V_d = \text{Reconstructed data volume}^{1} \\
V_s \text{ [TB]} &= 0.1 \cdot V_d \cdot S \cdot X = \text{Simulated data volume}^{2} \\
C_r \text{ [}10^6\text{ hours]} &= 0.024 \cdot T \cdot P = \text{CPU for real data}^{3} \\
C_s \text{ [}10^6\text{ hours]} &= 2 \cdot C_r \cdot S \cdot X = \text{CPU for simulation}^{4}
\end{aligned}
$$

1. The factor of 0.4 accounts for both decoded and DST data.
2. The factor of 0.1 is the size reduction from raw data to DST.
3. Assuming 1 Hz/core, based on current software and nominal luminosity.
4. The factor of 2 accounts for GEANT in addition to reconstruction.

3 2024 Submission Form Values

$$
\begin{aligned}
\text{Silo/Mass Storage} &= V_d + V_r + V_s \\
\text{Amount of Simulated Data Expected} &= V_s \\
\text{Amount of Raw Data Expected} &= V_d \\
\text{Amount of Processed Data Expected} &= V_r \\
\text{Online Storage Disk Required}^{5} &= 0.1 \cdot V_r \\
\text{Imported Data}^{6} &= V_s \\
\text{Exported Data}^{7} &= V_r \\
\text{Simulation Requirements}^{8} &= C_s \\
\text{Production}^{9} &= C_r
\end{aligned}
$$

5. This value should be increased for functionality not currently in standard use.
6. Ingress includes only simulations.
7. Egress includes only final data for physics analysis.
8. CPU resources for simulations are almost entirely offsite, not at JLab.
9. CPU hours are for JLab's current compute farm, not the requested SPEC CINT2000.