Next Generation Data-Intensive Analysis Framework for High Performance Computing Systems

Jeff Porter (NSD/NERSC)

Mateusz Ploskon (NSD)

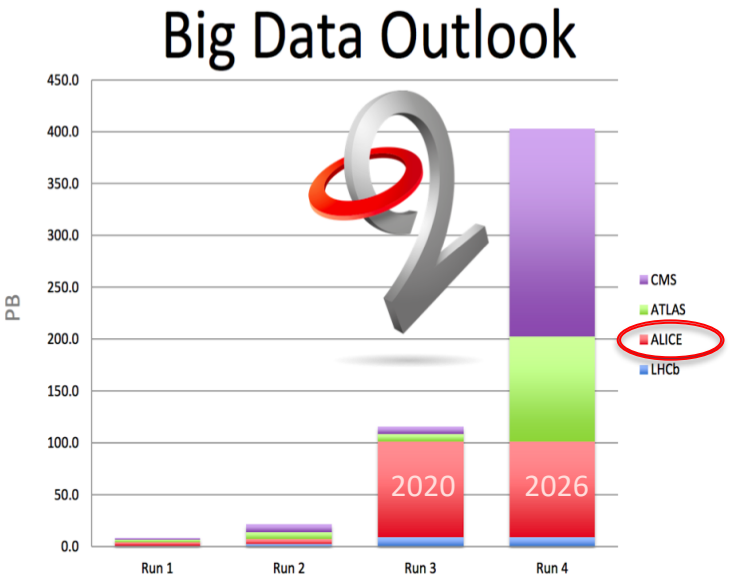
Lisa Gerhardt (NERSC)

Data Analysis Computing Challenge in High Energy Nuclear Physics Experiments

- **Example – LHC High Luminosity Era for ALICE**
  - In 2020, Detector, DAQ and beam upgrades boost data rates by over 100x
  - Major ALICE O^2 project exists to manage real-time data volume reduction
  - Retain the 100x boost in event sample
    - 100 million PbPb events in 2015 → 25 billion PbPb events in 2020

- **Analysis already uses large share of comp. resources**
  - Will also require a new processing model

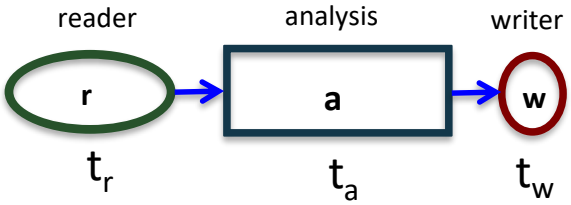
- **HPC systems becoming more "data friendly"**
  - NERSC Cori, 27 PB high bandwidth file system
    - Data Phase 1 2015 → 50k cores with 4GB/core
    - Phase 2 2016 → 600k cores with 1.2GB/core
  - Next Generation system at NERSC will come in ~2019



➢ **Project Goal: Build framework for efficient, high-throughput data analysis on many-core HPC**

Analysis of HENP event-based experiment data

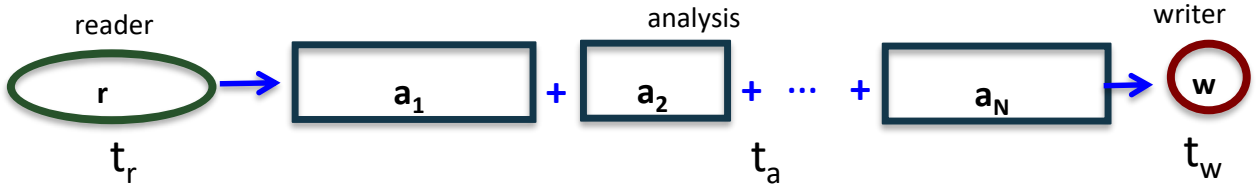
- **HENP Data-intensive workflow characterized by low CPU efficiency**
  - Large fraction of time spent in I/O relative to CPU



CPU efficiency = cpu time/wall time

$=\sim \quad t_a / (t_r + t_a + t_w)$

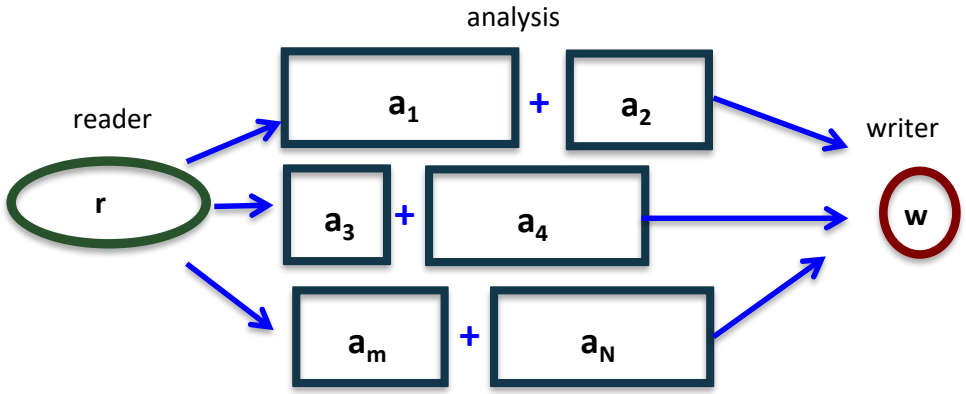
- **Analysis trains increase CPU efficiency by increasing payload**
  - Chain many analysis together in serial – read 1 for N analyses
  - Used for many years in several contexts: PHENIX, ALICE@GSI, ALICE Grid



- **Problem: increased memory usage & turn-around times**
  - Extreme challenge with large event samples from high-luminosity era

Proposal: Develop a High-Performance Parallel Data Analysis Framework

- **Develop a framework to parallelize trains into analysis-segments**
  - Reduced memory footprint per process & overall turn-around time
  - Retain read once for N analyses, fed by intelligent I/O + messaging service
  - Complexity is in the framework, not in the user analysis



- **R&D effort**
  - Develop intelligent reader/writers for high data ingest/digest rates
  - Optimized serialization/de-serialization, message passing, train configuration

- **ALICE & STAR Computing teams are extremely supportive**

LDRD Proposal

- **50% Physics PD STAR, 50% Physics PD ALICE**
  - Experiment independent framework
    - Customized reader/writer codes (e.g. ROOT-based)
    - Event/result message passing services
  - Evaluate & optimize processing chains within STAR & ALICE
  - Feedback common bottlenecks to Experiment Computing teams
- **10% Mateusz Ploskon NSD & 10% Jeff Porter NERSC**
  - Management and PI duties
  - Interface with ALICE and STAR computing teams
  - Guide effort with respect to technology options
- **20% of Lisa Gerhardt, NERSC**
  - Workflow optimization, Batch, MPI
  - Systems support at NERSC
- **Total costs**
  - Pre-G&A $216k/year for effort and travel for 2 years
  - No hardware costs expected