Modeling Analysis Computations and End-to-end Simulation-analysis Workflows

Preeti Malakar[1], Gagan Agrawal[2], Tekin Bicer[1], Venkatram Vishwanath[1], Todd Munson[1], Rajkumar Kettimuthu[1], Ian Foster[1]
[1]Argonne National Laboratory, [2]Ohio State University

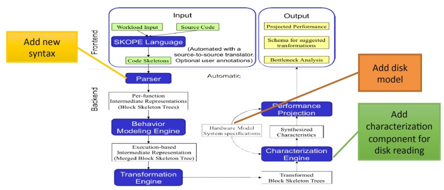
MOTIVATION

- Simulations and experiments are costly to perform, in terms of procuring the required resources, and job wait times.
- Analysis of petabytes of data from simulations and experiments require strategically utilizing the analysis resources.
- Modeling the performance of analysis, simulation-analysis workflows and experiment-analysis workflows helps scientists plan their simulation and experiment campaigns.
- Scientific experiments at synchrotron light sources are configuration-sensitive and depend on many parameters such as the number of projections and dose exposure time. Timely analysis of collected data can help deciding on configuration parameters and lead to more accurate data acquisition.
- Available compute resources at light sources are typically insufficient for analyzing the generated data, thus we require performance models and software tools to perform efficient analysis on distributed HPC resources.

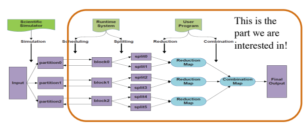
CONTRIBUTIONS

- Developed
  - Performance models for mapReduce-like in situ data analysis jobs
  - In situ and co-analysis execution workflow optimization models for scientific applications
  - Performance models for light source experimental data analysis workflow

MODELING IN SITU MAPREDUCE-LIKE ANALYSIS JOBS



SKOPE Extentions: Modeling Disk-Based and In-Situ Analytics



Smart: A MapReduce-Like Framework for In-Situ Scientific Analytics

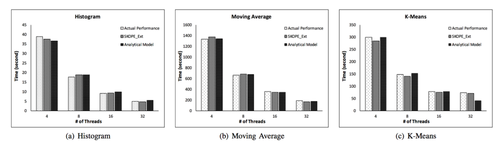
Model for memory access time of 2-level cache hierarchy: $T_{mem} = N_{mem} * (t_{l1-hit} + P_{l1-miss} * (t_{l2-hit} + P_{l2-miss} * L_{mem}$

For applications used in our experiments, the cache miss rate is dependent on dataset size (D), data access stride(s), cache capacity (C1 and C2 for L1 and L2 respectively) and block size (B1 and B2 for L1 and L2 respectively)

Scenario	Array Size	Stride	Frequency of L1 Misses	Miss Rate of L1	Frequency of L2 Misses	Miss Rate of L2
1	$1 \leq D \leq C_1$	$1 \leq s \leq D/2$	no misses	0	no misses	0
2.a	$C_1 < D \leq C_2$	$1 \leq s \leq B_1$	one miss for every B_1/s elements	s/B_1	no misses	0
2.b	$C_1 < D \leq C_2$	$B_1 < s \leq B_2$	one miss for every element	1	no misses	0
2.c	$C_1 < D \leq C_2$	$B_2 < s \leq D/2$	one miss for every element	1	no misses	0
3.a	$C_2 < D$	$1 \leq s \leq B_1$	one miss for every B_1/s elements	s/B_1	one miss for every B_2/s elements	B_1/B_2
3.b	$C_2 < D$	$B_1 < s \leq B_2$	one miss for every element	1	one miss for every B_2/s elements	s/B_2
3.c	$C_2 < D$	$B_2 < s \leq D/2$	one miss for every element	1	one miss for every element	1

Cache Performance Model for a Two-Level Cache Hierarchy

By defining parameters in the hardware model in SKOPE, and collecting memory access information such as read, write, access stride from code skeletons, we can utilize SKOPE to predict cache performance.

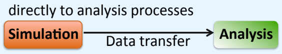


Predicting Scalability of Smart (conducted on the TACC Stampede cluster using 4 nodes)

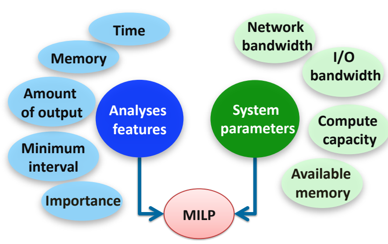
MODELING END-TO-END SIMULATION-ANALYSIS WORKFLOW

Mode 1: In situ analysis
Simulation and analysis alternates at a desired interval

Mode 2: Local co-analysis
Simulation processes communicate data directly to analysis processes



Mode 3: Remote co-analysis
Simulation communicates data to analysis via some form of storage



Mixed integer linear programming (MILP) based model incorporates various system and application parameters

Motivation
Derive maximum scientific insight from the simulations with minimum delay

Objective of Linear Program
Maximize number of different important analyses and their frequencies

Constraints
Throughput, available memory/storage, analysis specifications

Output
Analyses feasibility and their frequencies

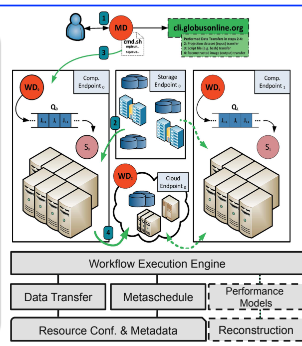
Experiments
Applications: LAMMPS, FLASH, miniMD
Systems: 32,768 cores on Mira Blue Gene/Q, 1080 cores on Edison Cray XC30

Results
100x faster than post hoc analysis
More than 85% utilization of allowed time for simulation-time analysis

MODELING LIGHT SOURCE DATA ANALYSIS WORKFLOW

Motivation
- Technological advances in x-ray sources and detectors enable increasingly complex experiments and rapid data acquisitions (e.g. 16GiB/s data generation rates)
- Timely analysis of this data is important
- Experimental facility users have limited time to perform experiments and collect data
- Near-real-time data processing is highly desirable to verify accuracy of data and determine proper configuration for experiments

Objectives
- Improving the parallelization of tomographic reconstruction jobs to minimize turnaround time
- Enabling utilization of geographically distributed resources
- Estimating execution times of different stages in workflows
- Data transfer, queue/idle time, computation time



Implementation
- A MapReduce-like middleware for tomographic reconstruction algorithm
- A workflow management system for light source data analysis jobs
- Models and methods for estimating execution time of workflow stages
- Estimation of data transfer rates between resources
- Bandwidth Delay Product + Initialization Cost
- Cluster queue time prediction
- Simulation of target cluster's job queue
- Modeling the performance of tomographic reconstruction jobs

Experiments
Applications: Iterative tomographic reconstruction algorithms
HPC Resources: Mira Blue Gene/Q (32K cores), Stampede , Gordon

Results
- Reconstruction times decreased from days to minutes (24-core workstation vs. 32K cores on Mira BG/Q)
- Estimation of end-to-end workflow execution with 2.1-23.3% error rate

CONCLUSIONS

- Modeled performance of map-reduce like in situ data analysis kernels
- Extended SKOPE performance modeling framework to model cache
- Developed models to propose the optimal set of analysis computations that can be performed within time and space constraints
- Proposed mixed-integer linear programs for formulating optimal in situ and co-analysis execution workflows
- Shorter turnaround time of experimental data analysis workflow help verification of collected data and steering experiments at light sources
- Developed performance models and software tools for execution of light source data analysis tasks on distributed HPC resources

REFERENCES

- "Smart: a MapReduce-like framework for in-situ scientific analytics", Y. Wang et al., SC 2015.
- "Optimal Scheduling of In-Situ Analysis for Large-Scale Scientific Simulations", P. Malakar et al., SC 2015.
- "Optimal Execution of Co-analysis for Large-scale Molecular Dynamics Simulations", P. Malakar et al., SC 2016.
- "Rapid Tomographic Image Reconstruction via Large-Scale Parallelization", T. Bicer et al., EuroPar 2015.
- "Optimization of Tomographic Reconstruction Workflows on Geographically Distributed Resources", T. Bicer et al., Journal of Synchrotron Radiation 2016.

Acknowledgements
Funding Source: DOE ASCR RAMSES: Robust Analytical Modeling for Science at Extreme Scales
Compute Resources: Argonne Leadership Computing Facility