EXPERTISE

models, EXperiments and high PERformance computing for Turbine mechanical Integrity and Structural dynamics in Europe

ISC 2020 DIGITAL
JUNE 22-25
#ISC20

Predictive Modeling Supported Collective I/O Auto-tuning

Ayse Bagbaba

Universität Stuttgart

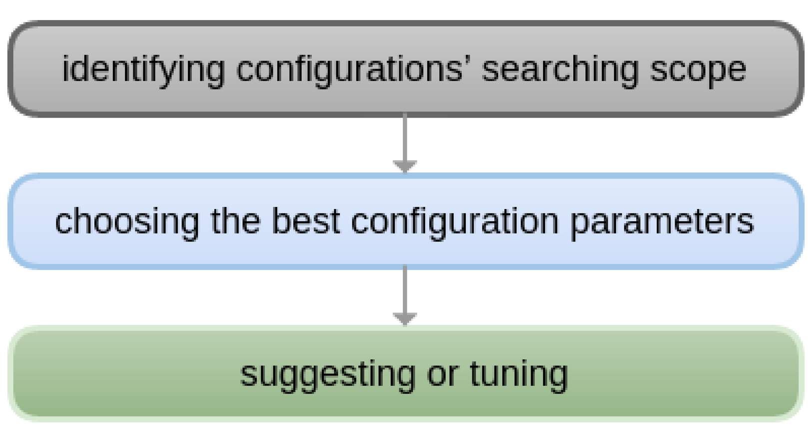
HLRS

Motivation and Objectives

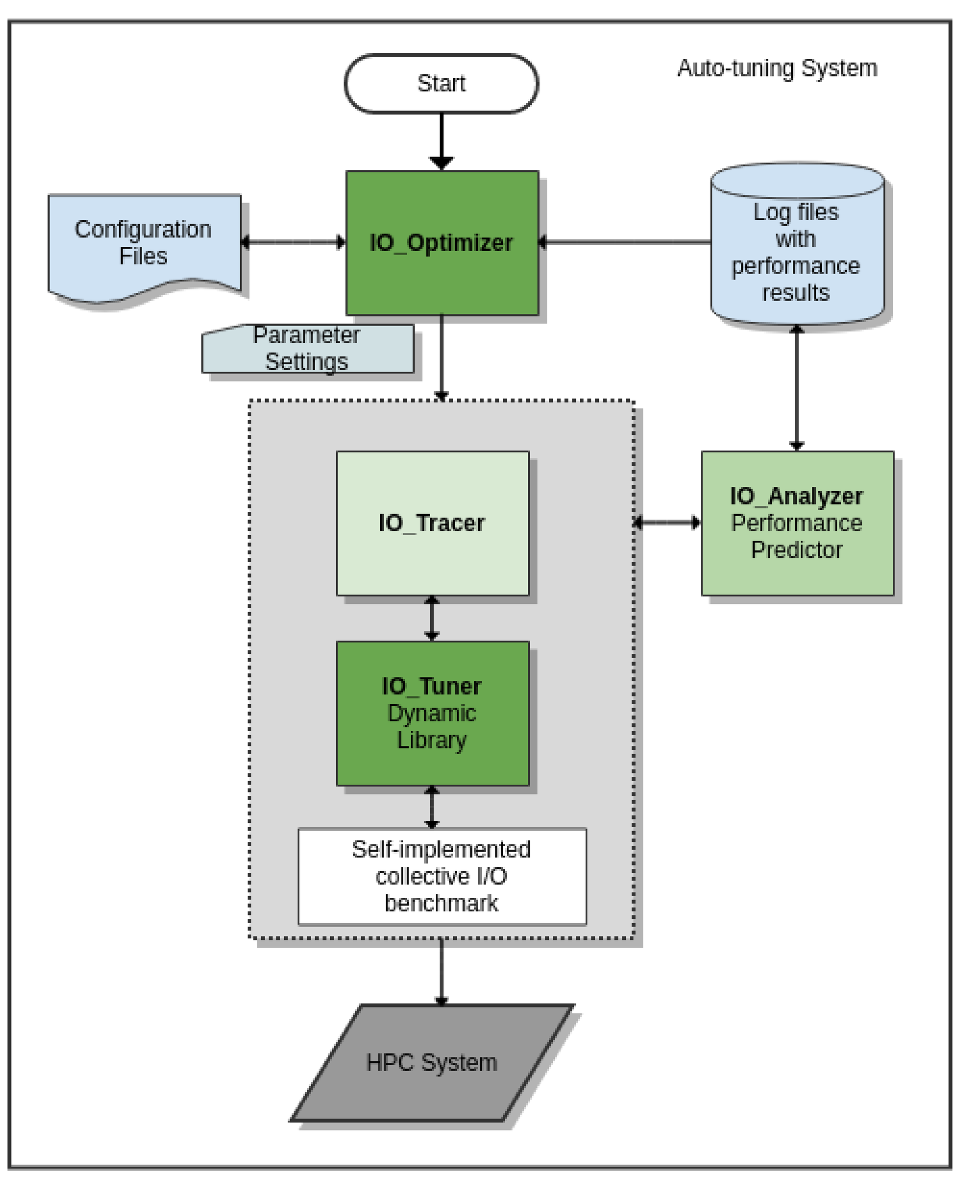
- **Motivation**:
- Effective parallel I/O is a nontrivial job due to the complex interdependencies between the layers of I/O stack;
  - the correct combination of a number of tunable parameters depends on diverse applications and HPC platforms,
  - engineers and scientists might not be capable of tuning their applications to the optimal level,
  - the default settings are often leading to poor I/O efficiency,
  - predicting the resulting performance is difficult.

- **Objectives:**
- An auto-tuning solution for optimizing collective I/O requests and providing system administrators or engineers the statistic information,
- Performance modeling including the architecture and software stack to analyze I/O requests.

Optimization Approach



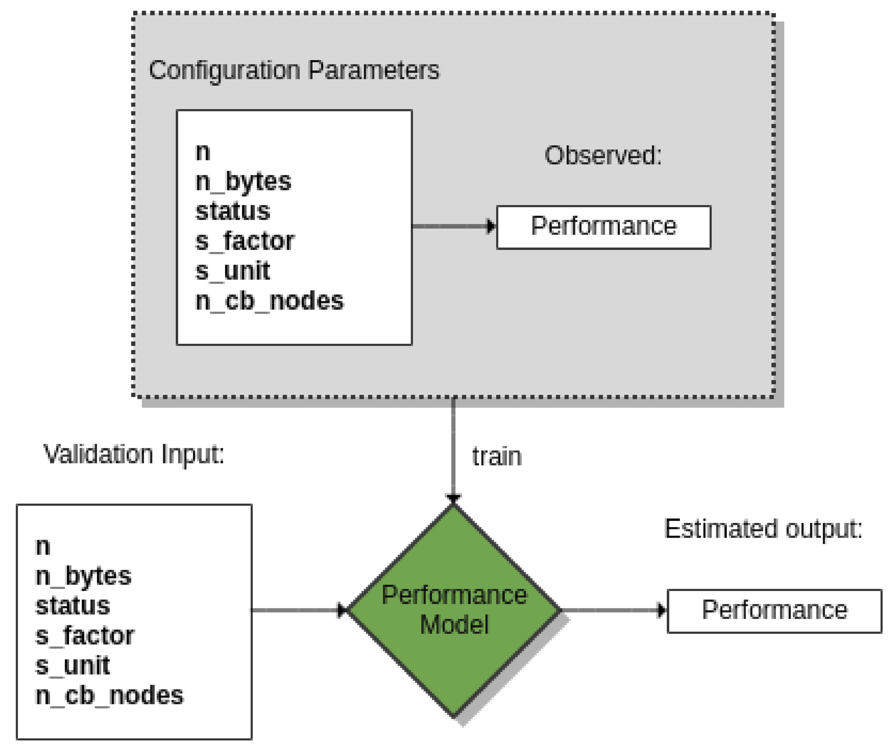
Auto-tuning



Overall Architecture of the Auto-tuning Approach with the following modules; *IO_Tracer, IO_Tuner, IO_Optimizer, IO_Analyzer.*

Predictive Modeling

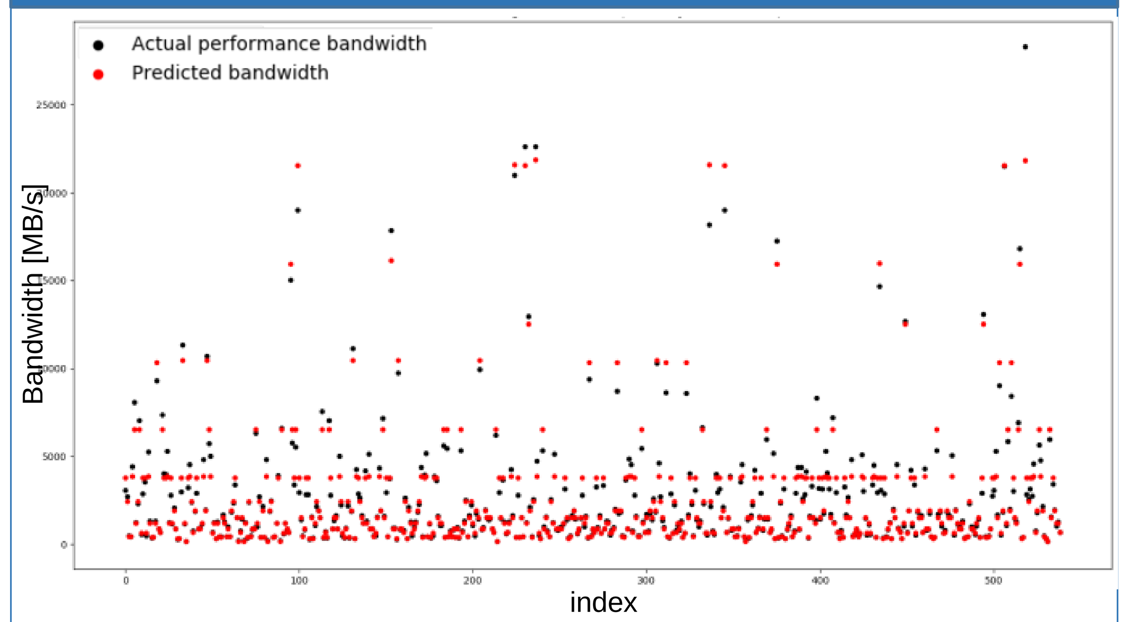
Consider a modeling approach that can model the I/O performance in terms of the application and file system configuration parameters such as *number of processes* (n), *number of bytes* (n_bytes), *collective buffering* (status), *number of collective buffering nodes* (n_cb_nodes), Lustre *striping factor* (s_factor), Lustre *striping unit* (s_unit) for "single file, collective clients" I/O access pattern.



$$\phi = f(\alpha, \zeta, \omega),$$

- α: a set of observable parameters that describe application characteristics (I/O pattern, I/O operation, benchmark)
- ζ: a set of observable parameters that describe file system and/or I/O characteristics (Lustre parameters)
- ω: uncontrolled non-observable parameters
- Aim: to understand the relationship between ϕ and the parameters (α, ζ). For a given set of input parameter values in (α, ζ), the function f should give a prediction.

Performance Model Evaluation



Random forest regression performance model with max depth = 4, Accuracy: 90.52 % on dataset including configuration parameters and achieved I/O bandwidth to be used in training and validation of the performance model (Training set size is 2153, test set size is 539). The performance model is extracted based on 100 iterations. Time taken to build model is 3.19 seconds.

max_depth	Prediction errors under different depths		
	Accuracy	MAE	RMSE
3	82.16 %	495.86	963.36
4	90.52 %	287.92	576.51
5	95.15 %	147.25	325.94
7	98.87 %	46.27	180.32
10	99.68 %	24.85	167.20

Prediction errors in MB/s for training sets under different depth of each tree in the random forest algorithm (*MAE*: Mean Absolute Error, *RMSE:* Root Mean Squared Error)

Experimental Results

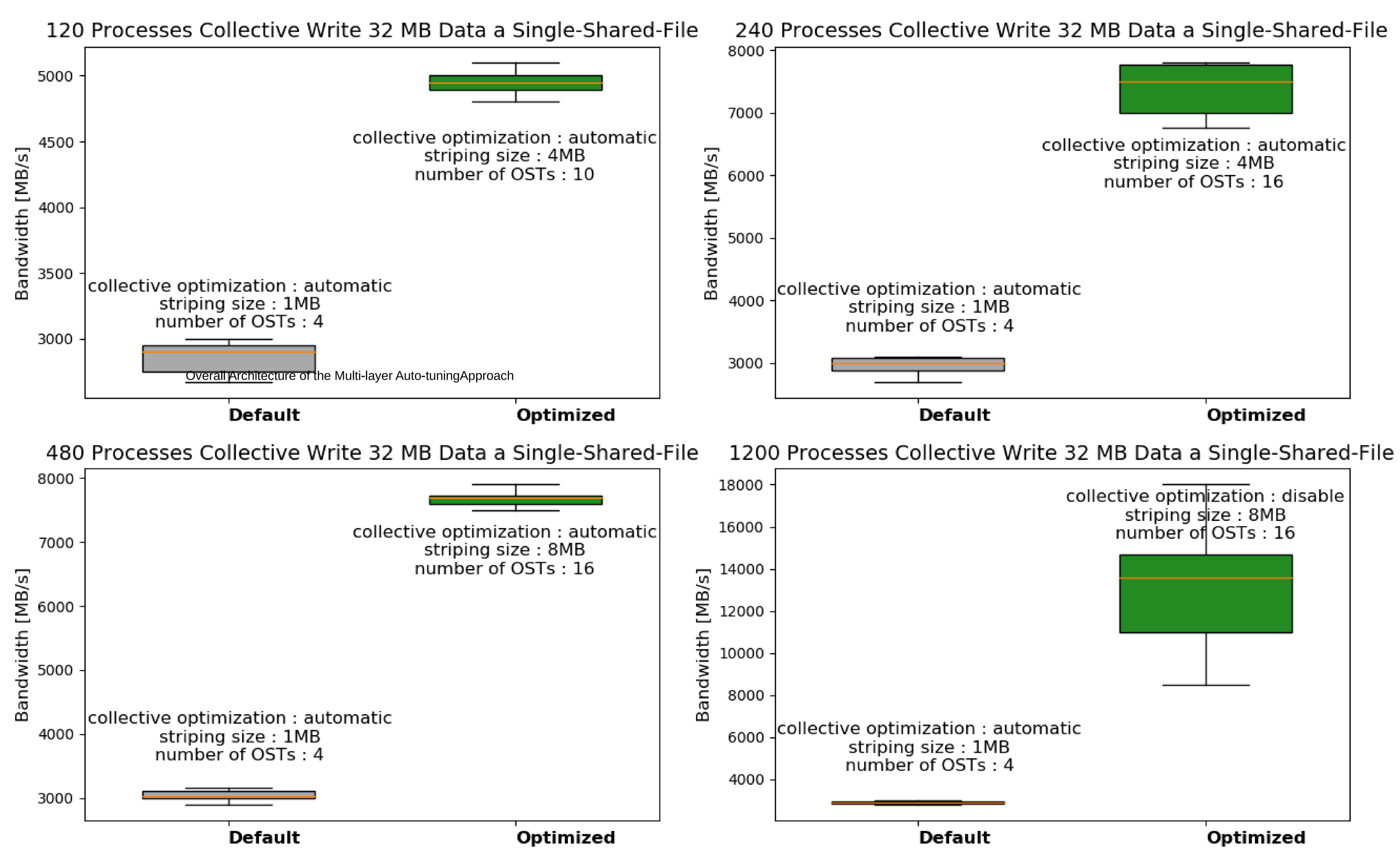


Table 1: Configurations' Searching Scope

Name	Value
n	24-1200
n_bytes	256 B - 196 MB
n_cb_nodes	1 - 16
s_factor	1 - 16
s_unit	1 MB - 32MB
status	automatic; disable; enable
IO pattern	collective

Table 2: Technical Details of Hazel Hen

Architecture	Cray XC40
Hardware	Intel Xeon E5-2680 v3 Cray Aries Network 7712 Compute nodes 90 Service nodes
File System	Lustre 7 MDTs 54 OSTs
Storage	Cray Sonexion 2000
Bandwidth	3.75 GB/s per OST

Conclusion and Future Work

Development of an predictive modeling supported auto-tuning solution to improve collective I/O performance;

- implemented upon widely used MPI-IO library,
- approachable for engineers/administrators with little knowledge of parallel I/O,
- compatible with MPI based scientific and engineering applications, portable to different HPC platforms,
- has a success varies between 50-130% at scale in the parametrization's I/O bandwidth gain,
- supports and informs the system administrators if performance anomaly is detected,
- will be tested on engineering applications in different professional areas.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 Framework Programme research and innovation programme under the Marie Sklodowska-Curie agreement No 721865.