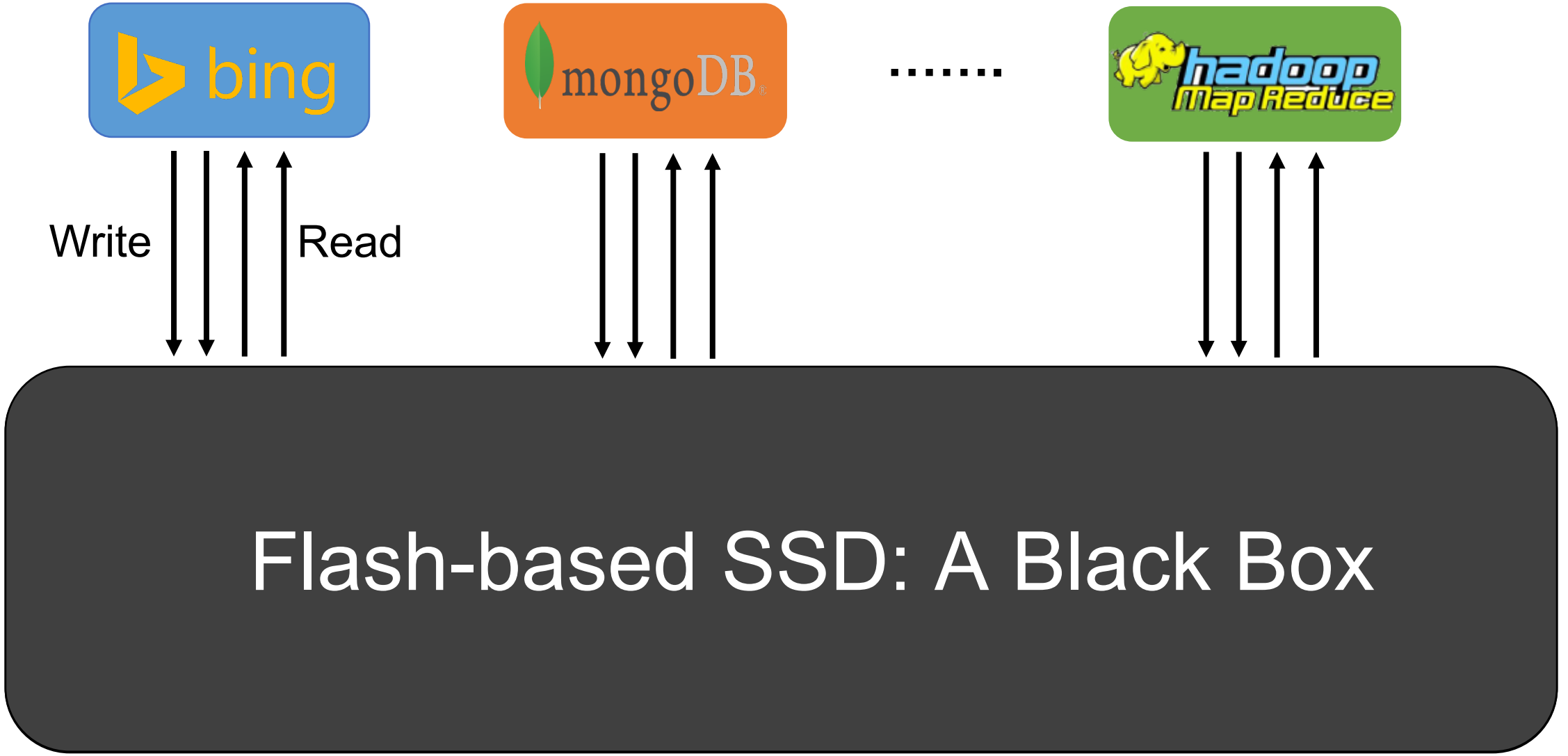
Learning to Drive Software-Defined Storage

Jian Huang Daixuan Li **Jinghan Sun**

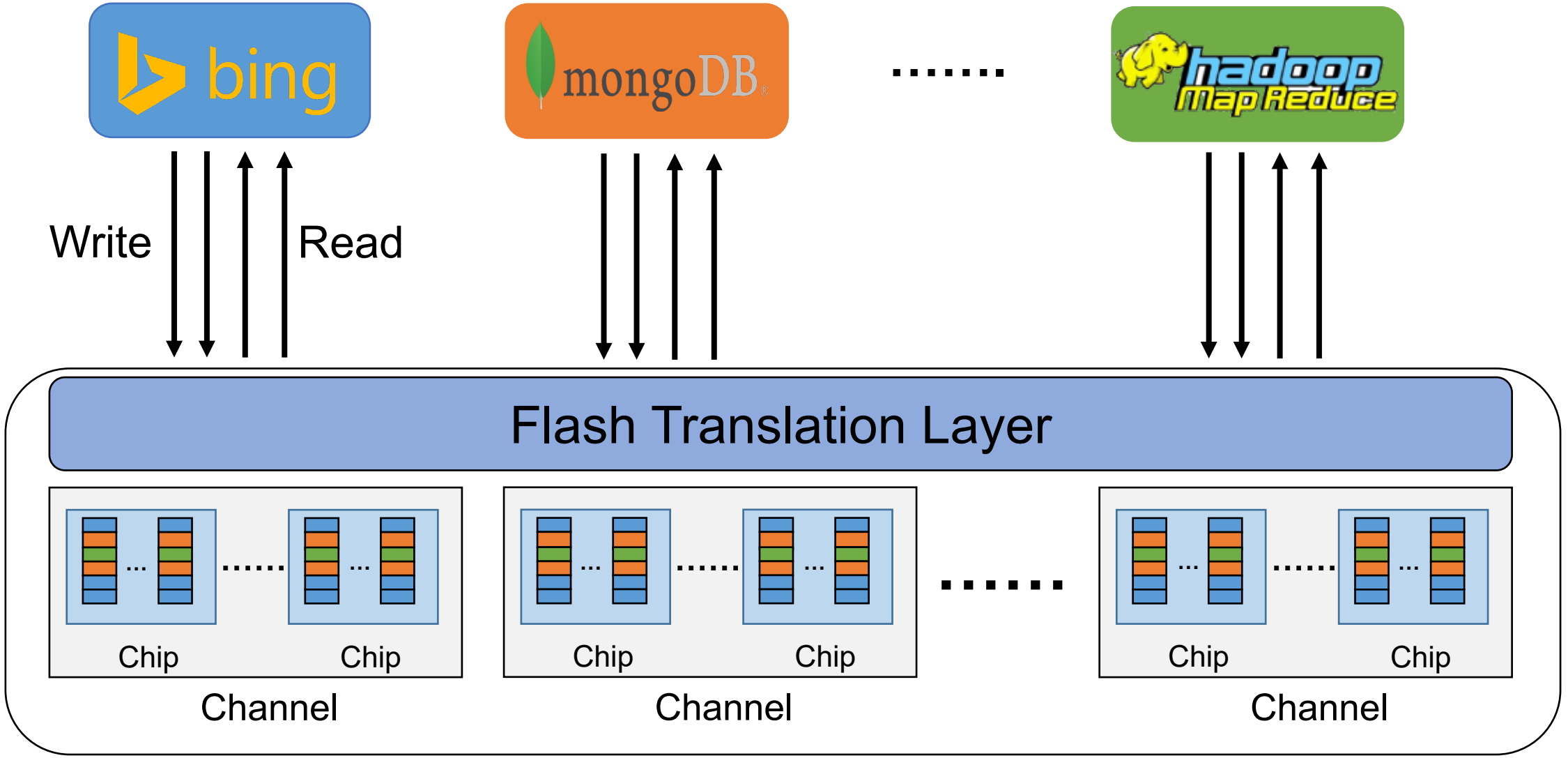
Systems Platform Research Group

UNIVERSITY OF
ILLINOIS
URBANA-CHAMPAIGN

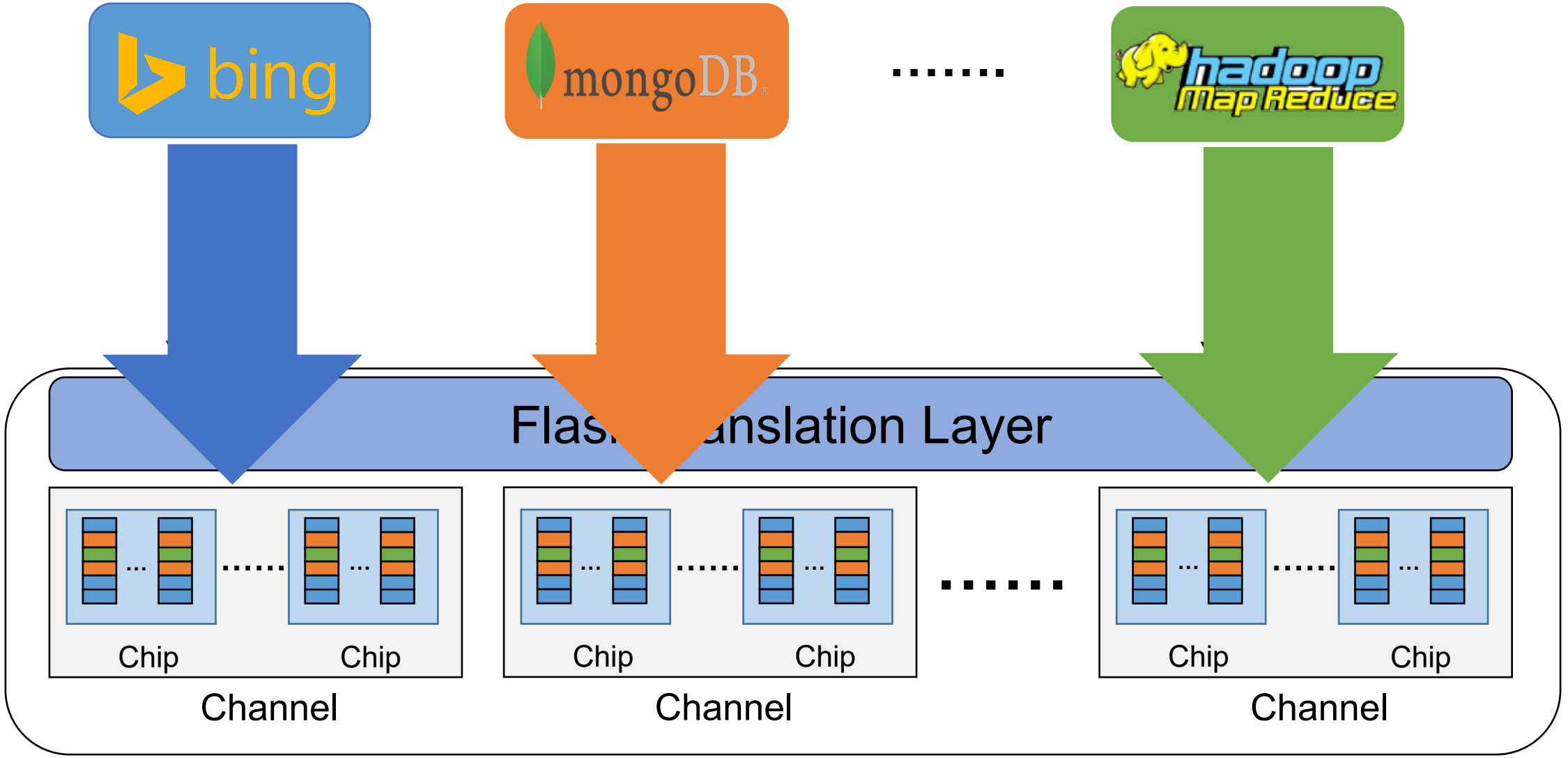
What is Software-Defined Storage?



What is Software-Defined Storage?



What is Software-Defined Storage?



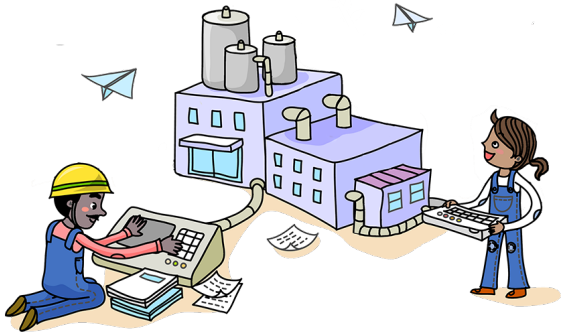
Benefits of Software-Defined Storage

Predictable Performance

+

Improved Storage Utilization

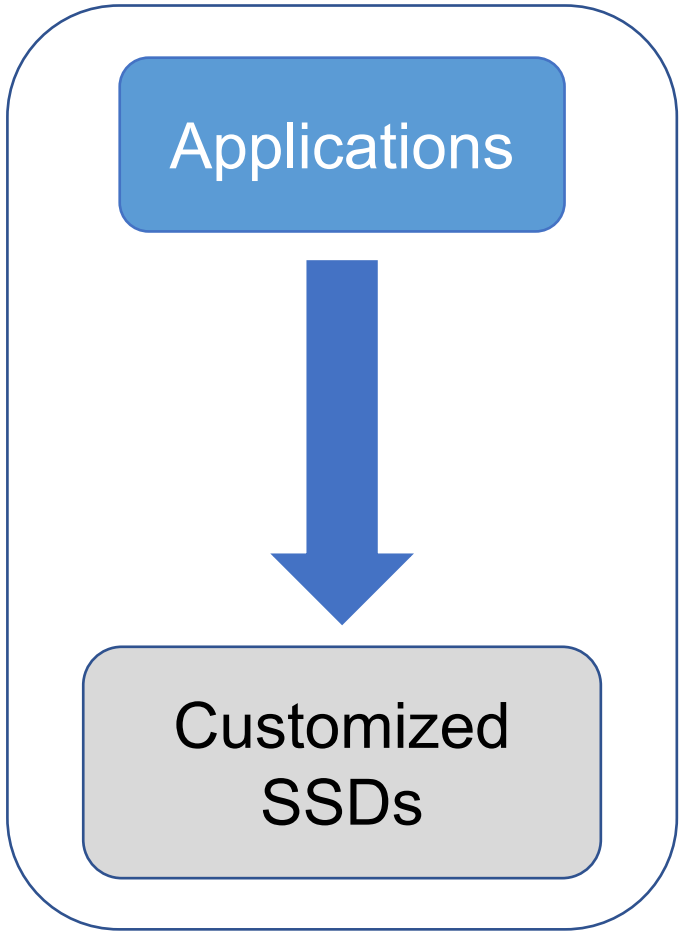
+



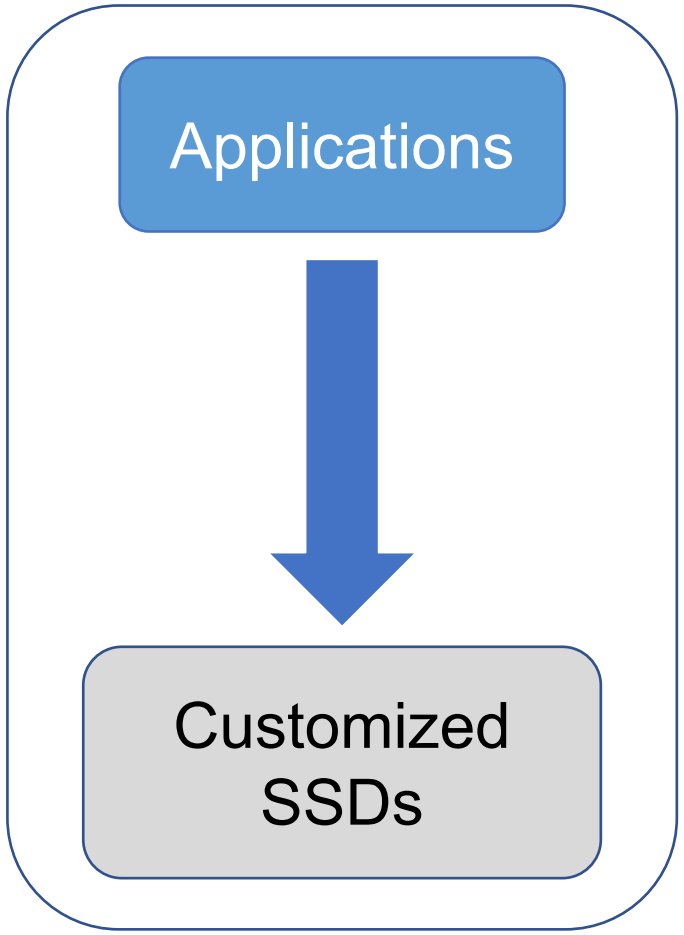
Mature Manufacturing Technologies

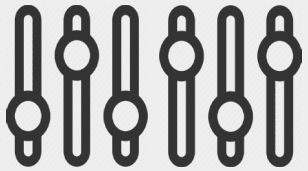
Software-Defined SSDs

Challenges of Building Software-Defined Storage



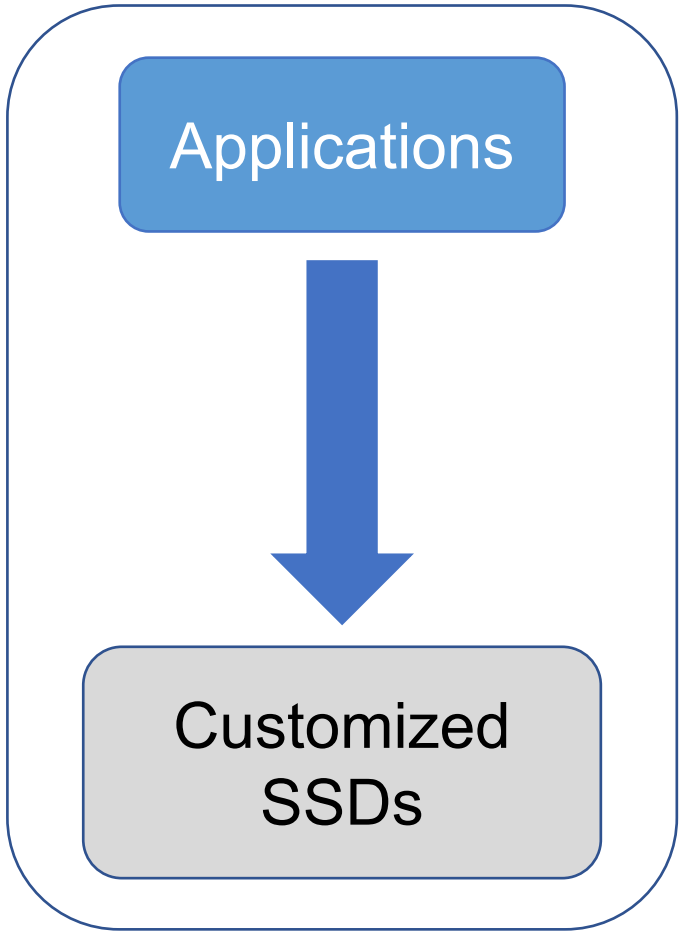
Challenges of Building Software-Defined Storage

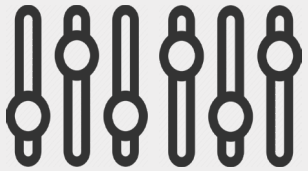




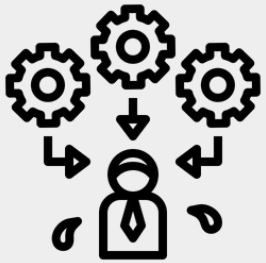
A large set of device parameters to tune

Challenges of Building Software-Defined Storage



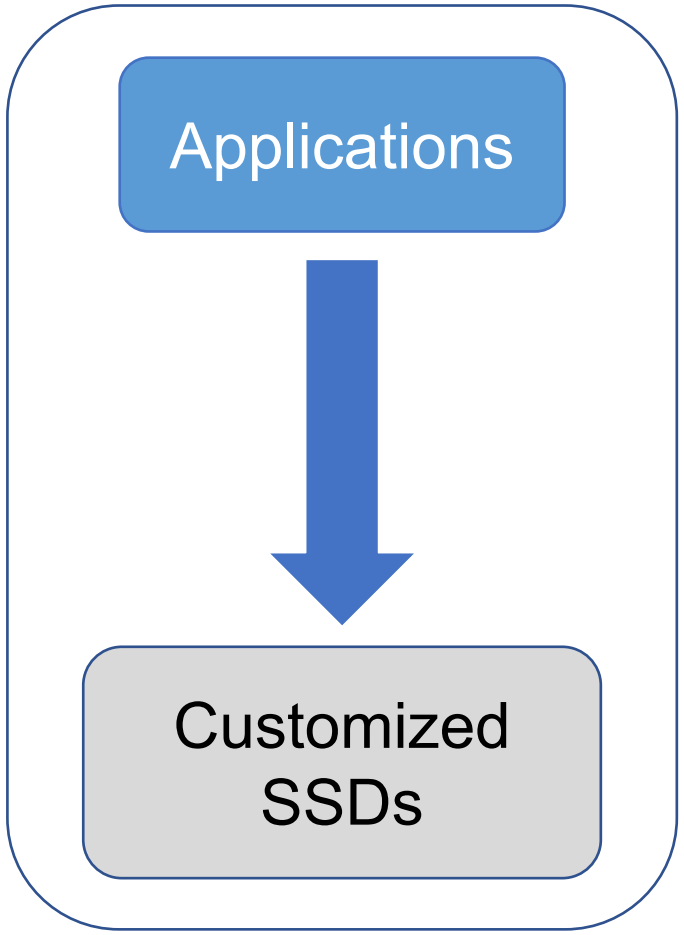


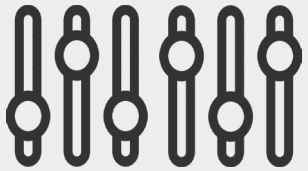
A large set of device parameters to tune



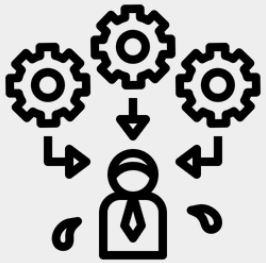
Hard to manually identify optimized device parameters for different applications

Challenges of Building Software-Defined Storage





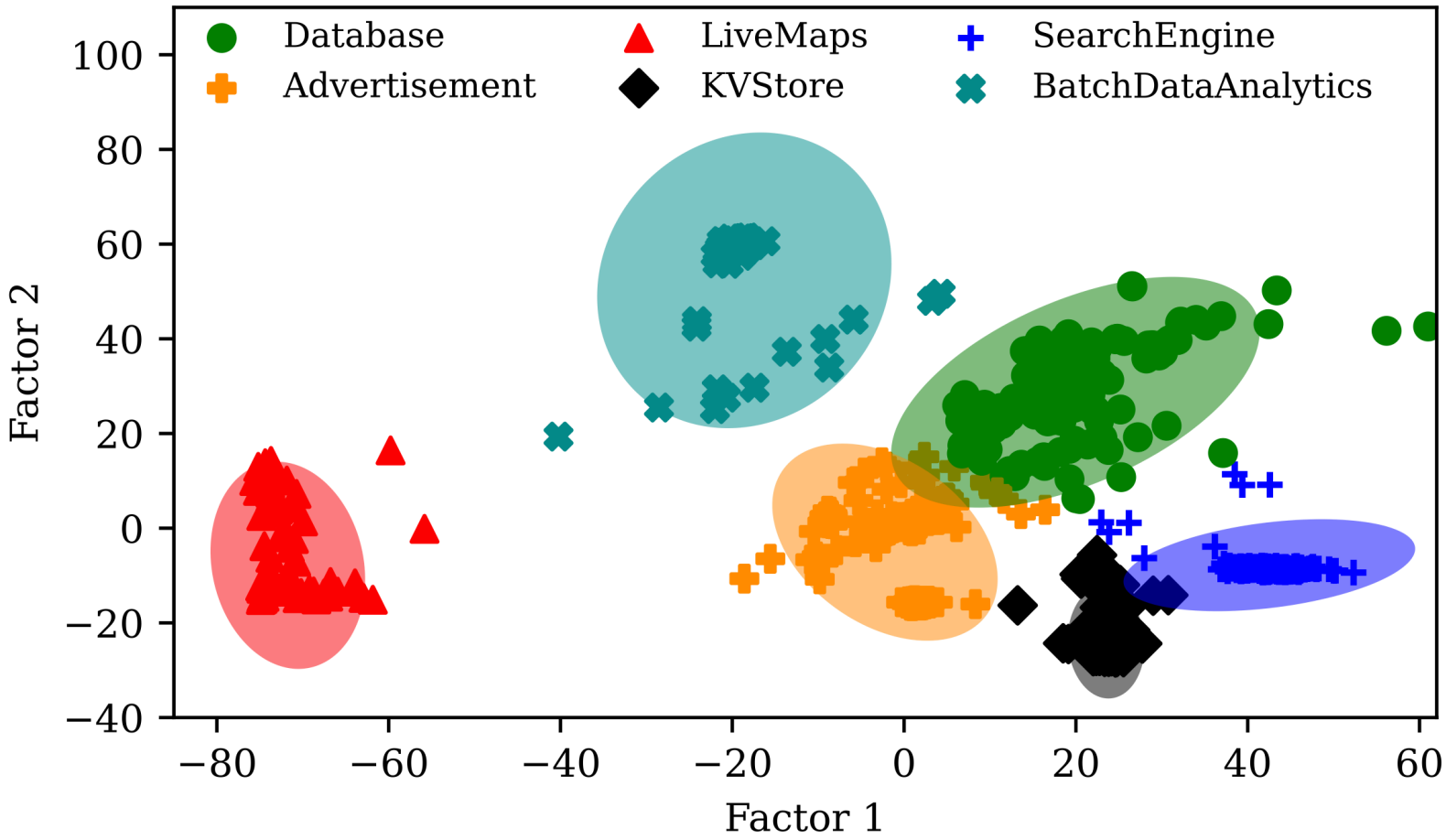
A large set of device parameters to tune



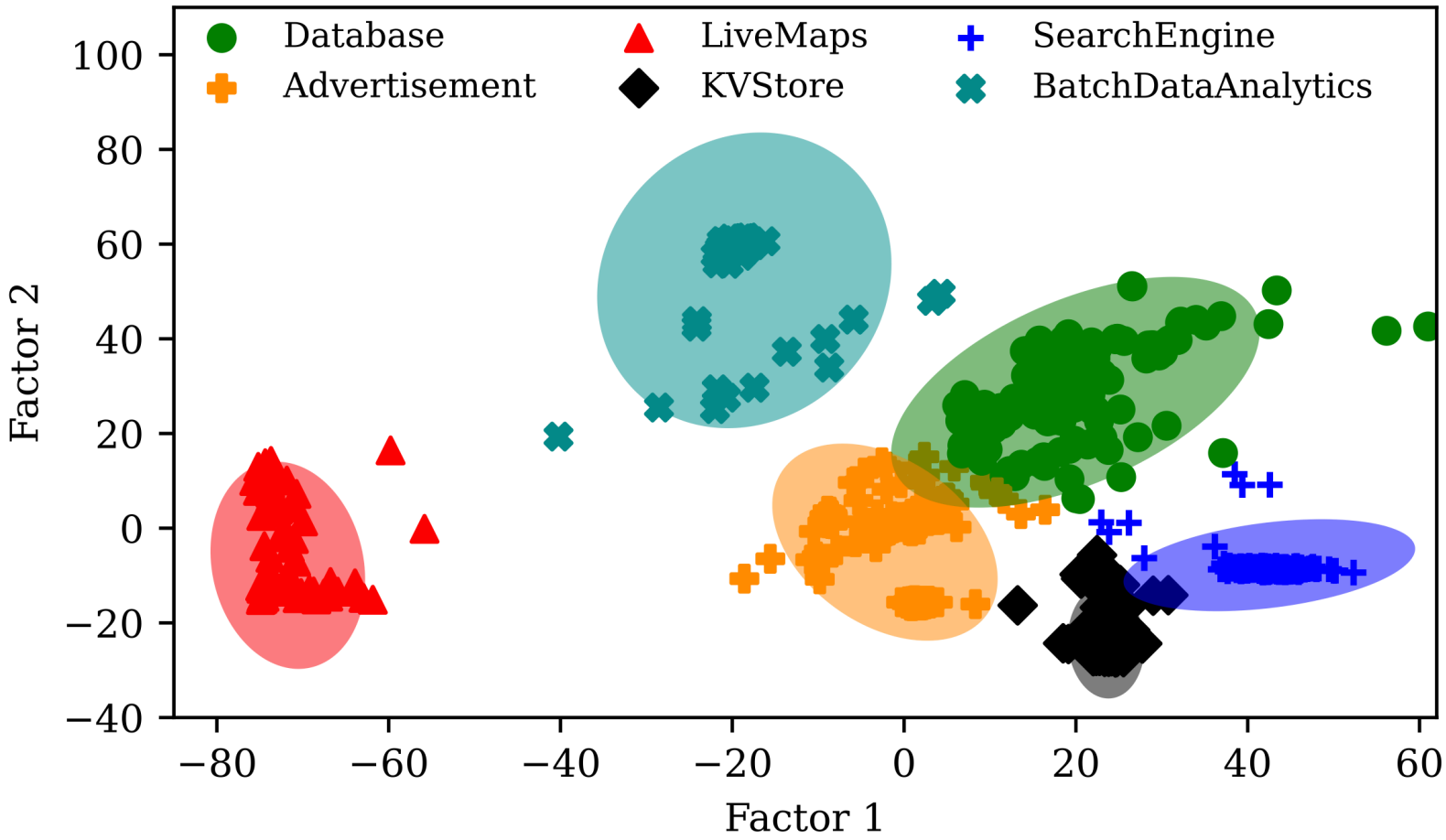
Hard to manually identify optimized device parameters for different applications

Traditional testing method has a very long tuning time

Learning to Drive Software-Defined Storage



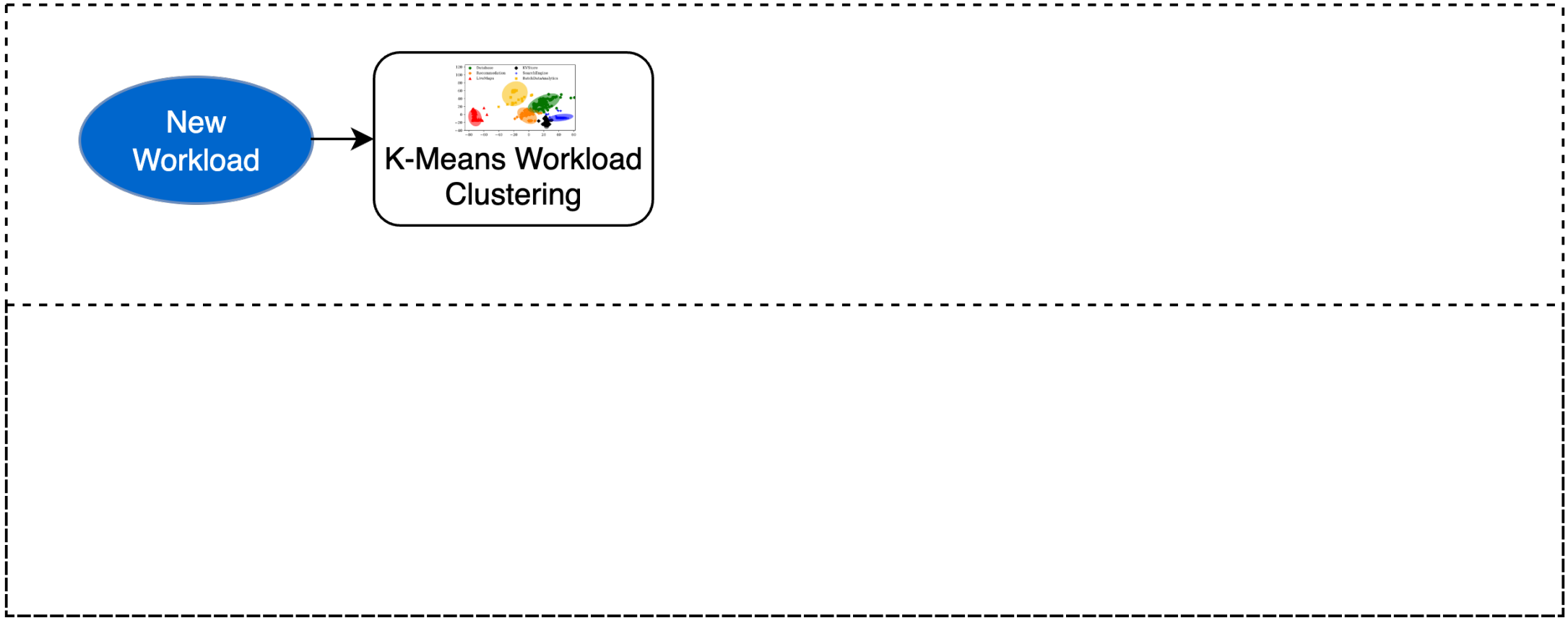
Learning to Drive Software-Defined Storage



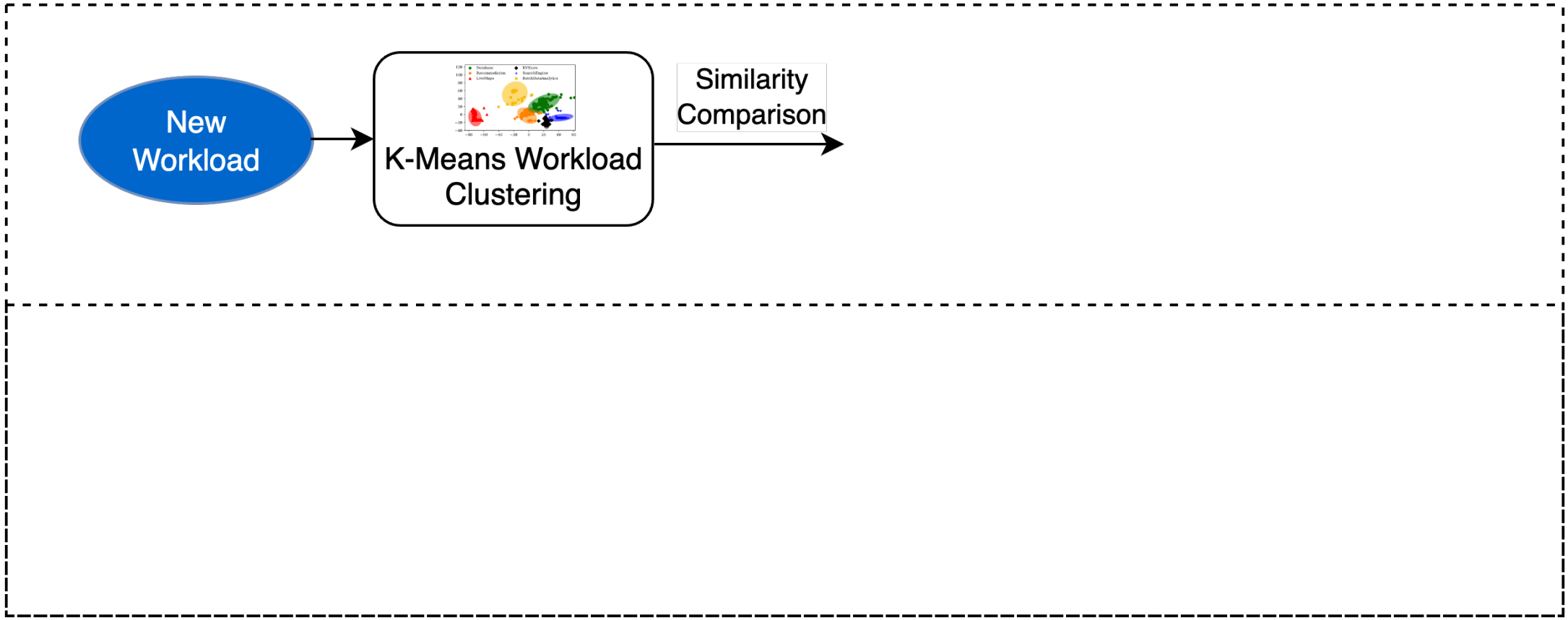
The learning-based apporach to developing software-defined storage is promising

The Workflow of Learning Optimized SSD Configurations

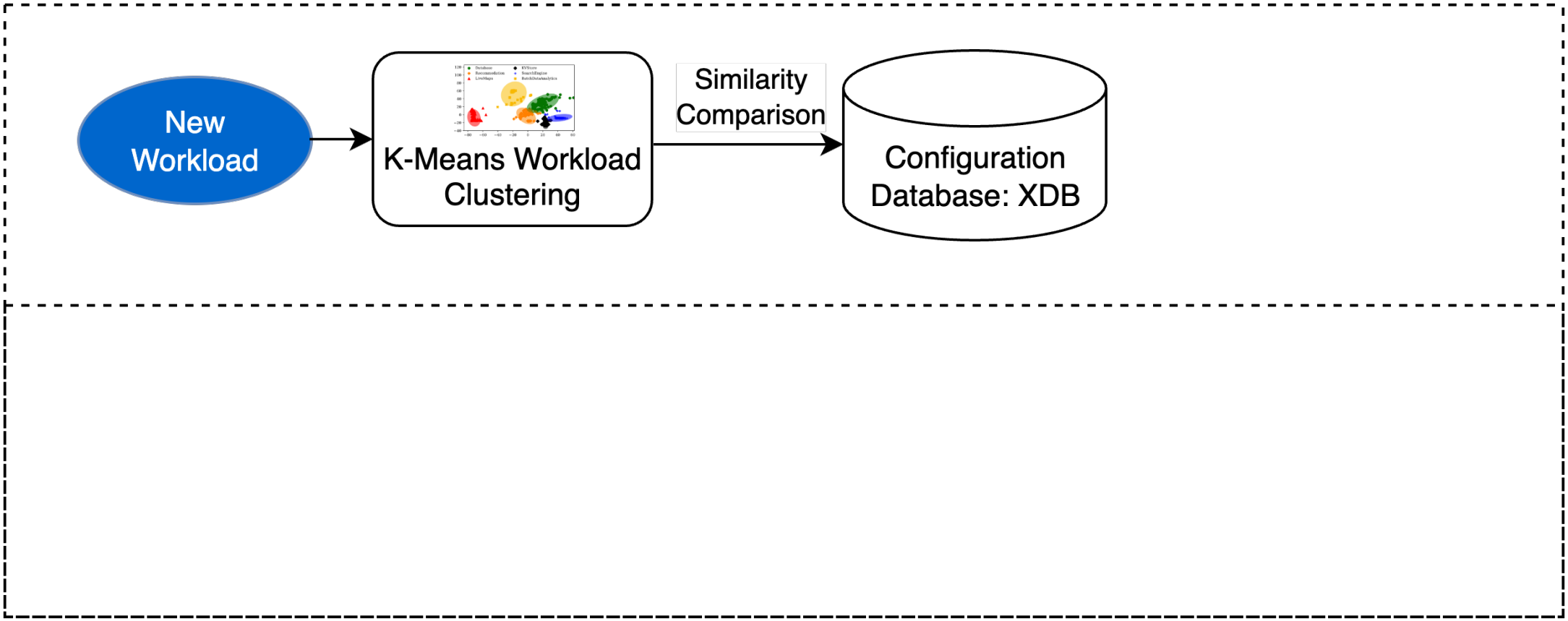
The Workflow of Learning Optimized SSD Configurations



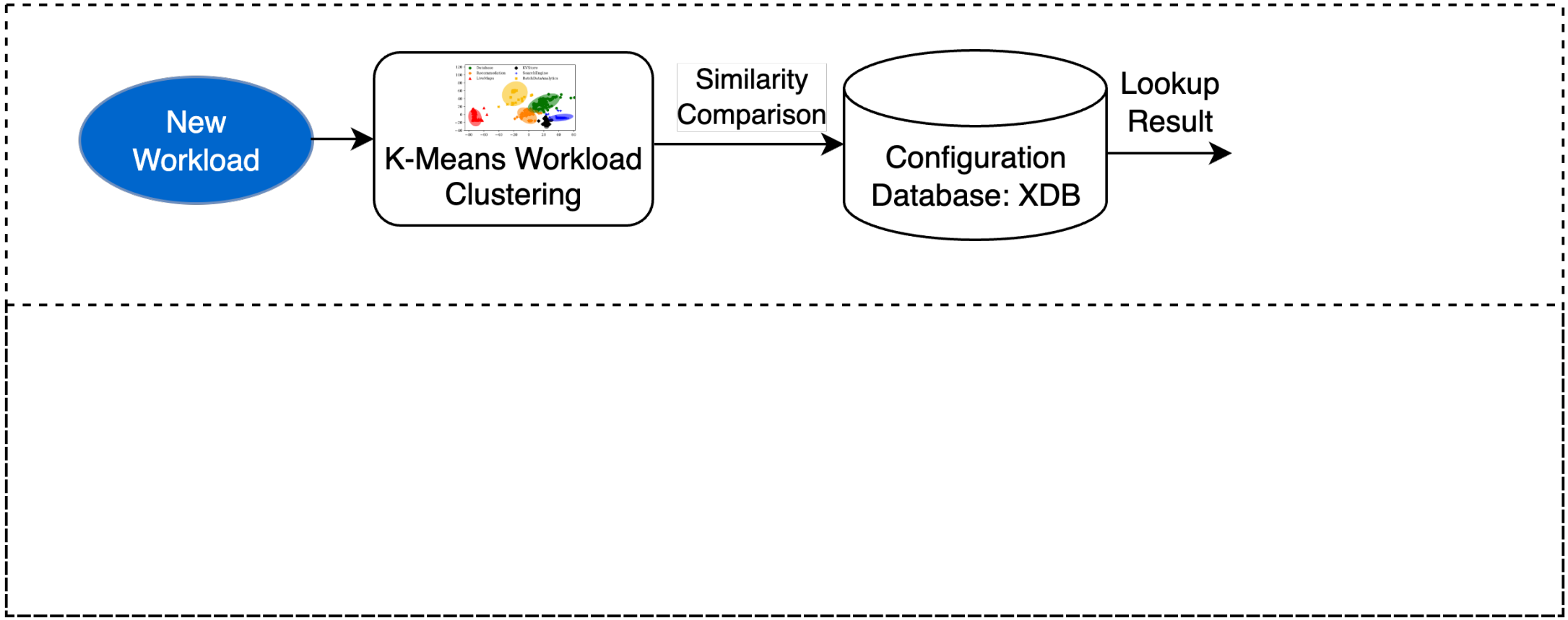
The Workflow of Learning Optimized SSD Configurations



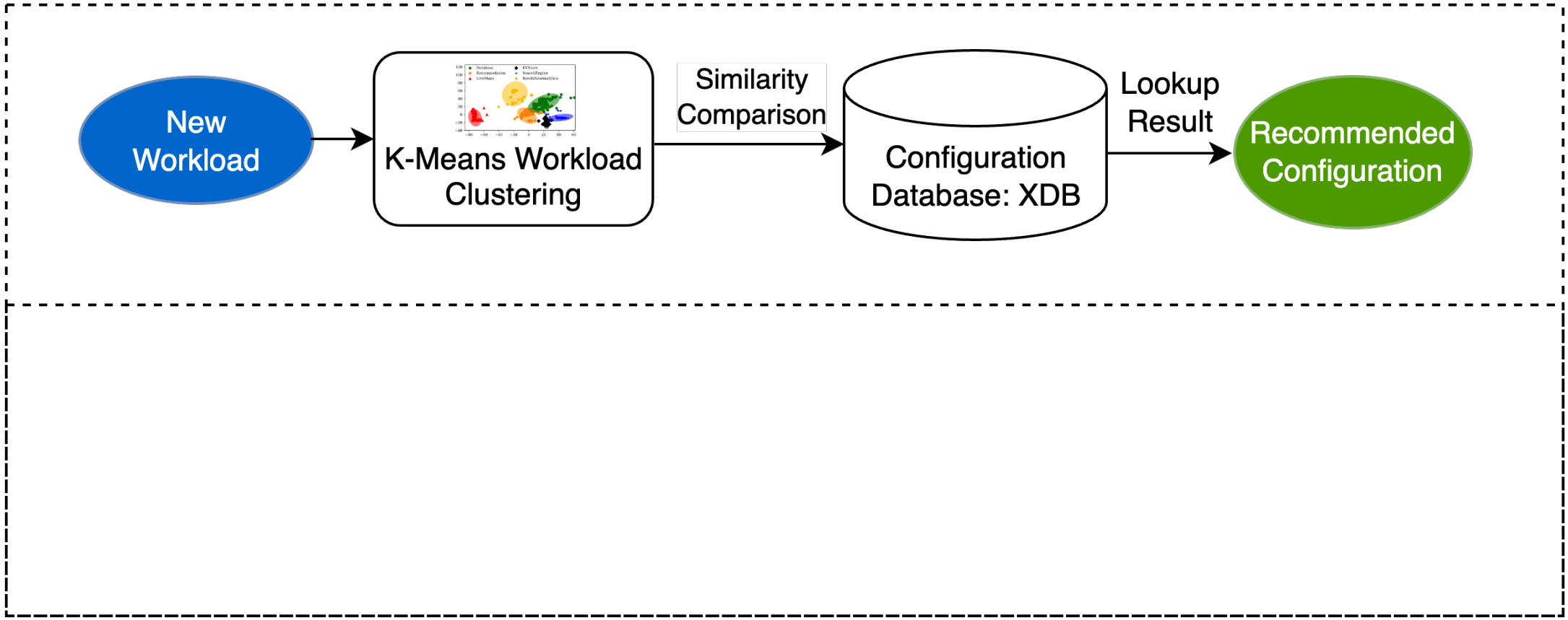
The Workflow of Learning Optimized SSD Configurations



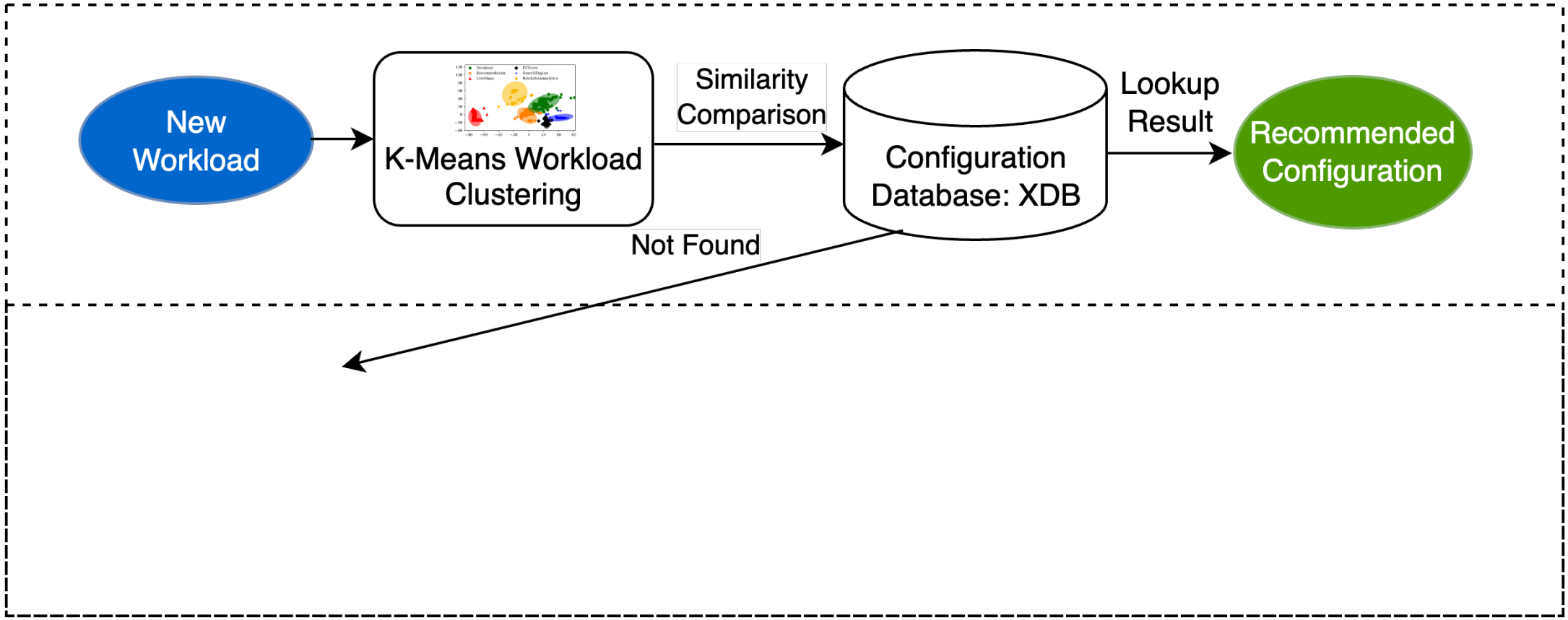
The Workflow of Learning Optimized SSD Configurations



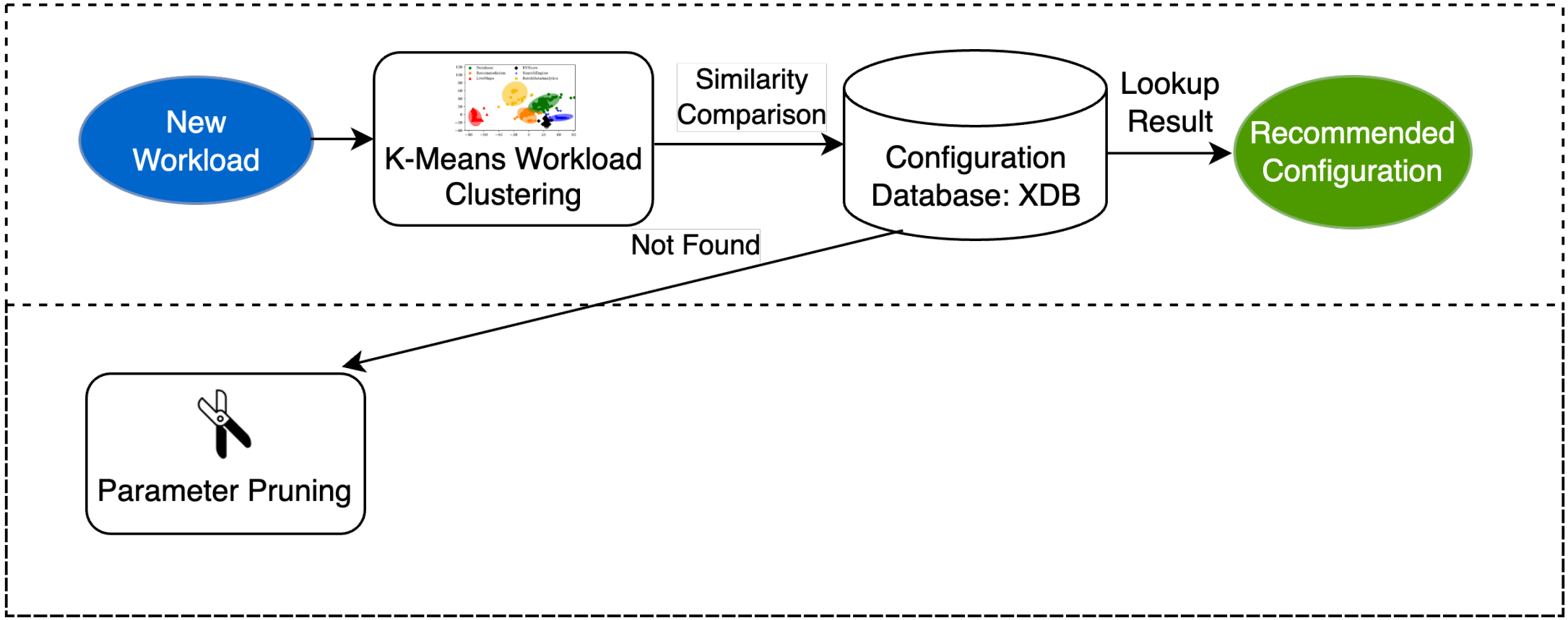
The Workflow of Learning Optimized SSD Configurations



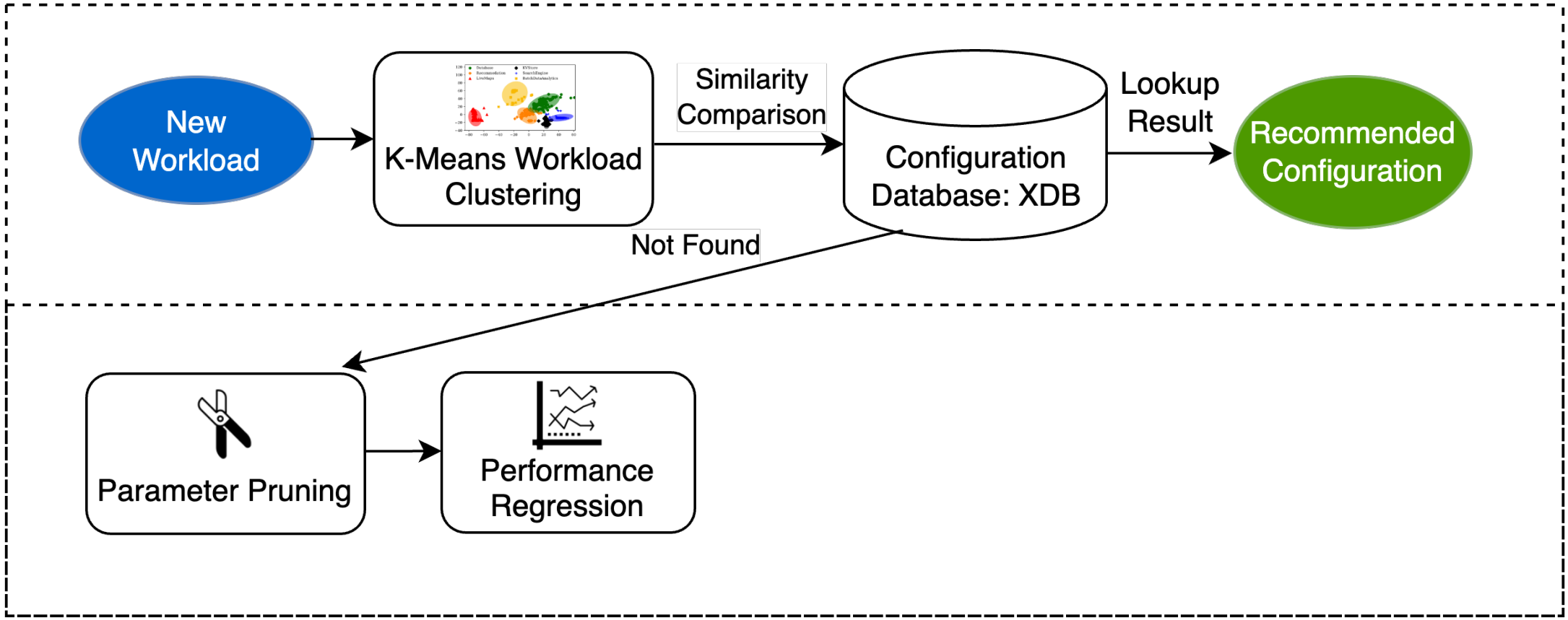
The Workflow of Learning Optimized SSD Configurations



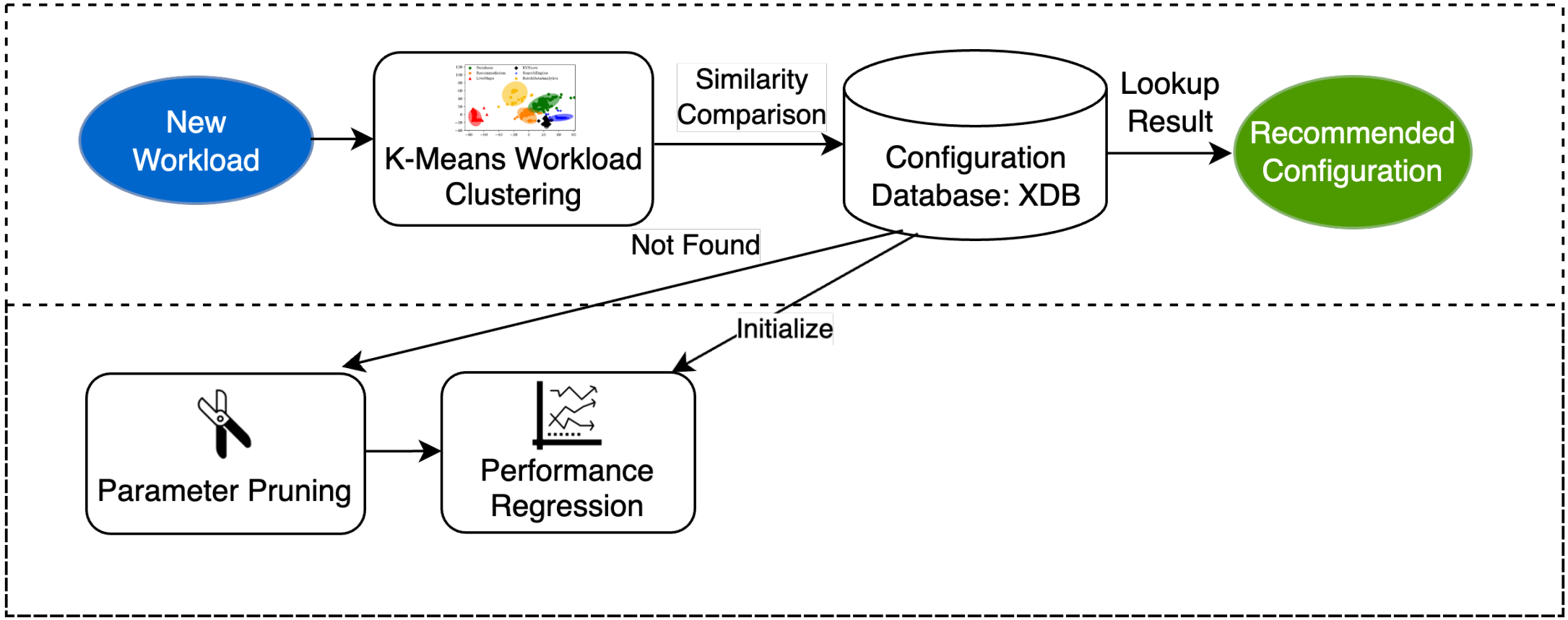
The Workflow of Learning Optimized SSD Configurations



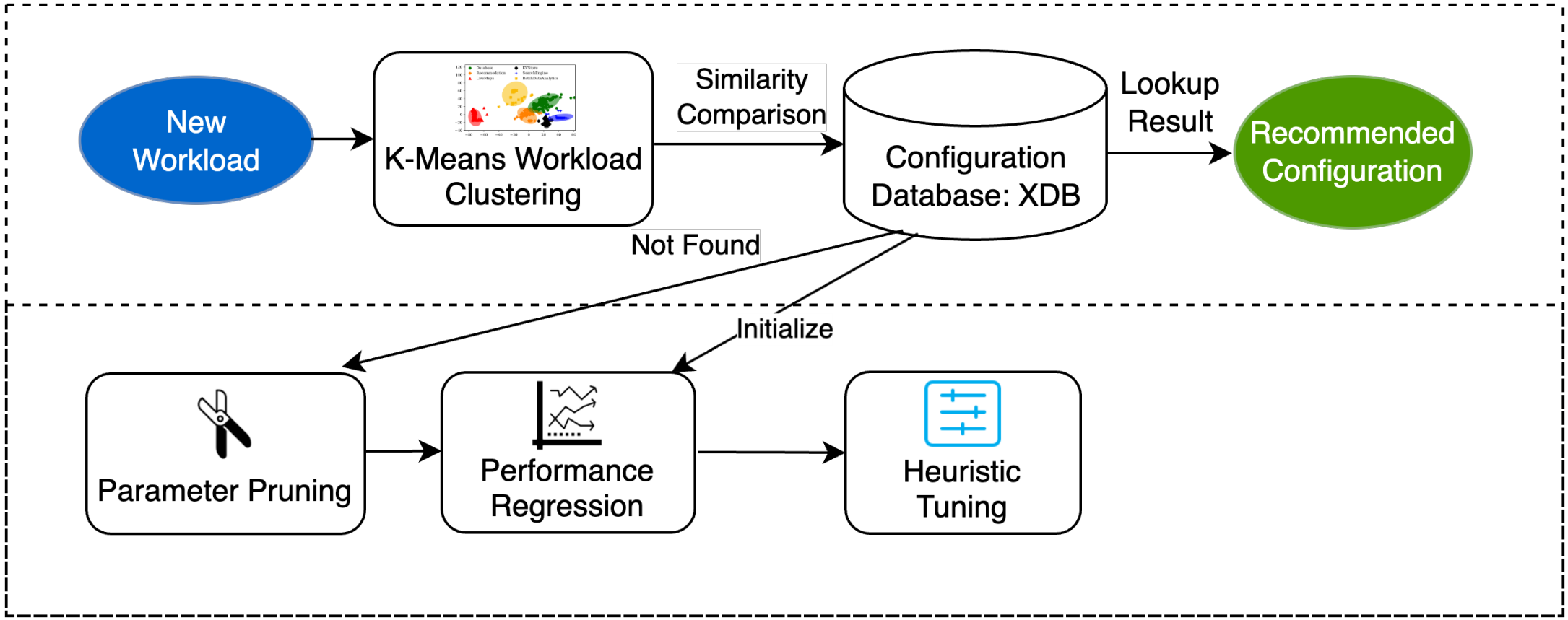
The Workflow of Learning Optimized SSD Configurations



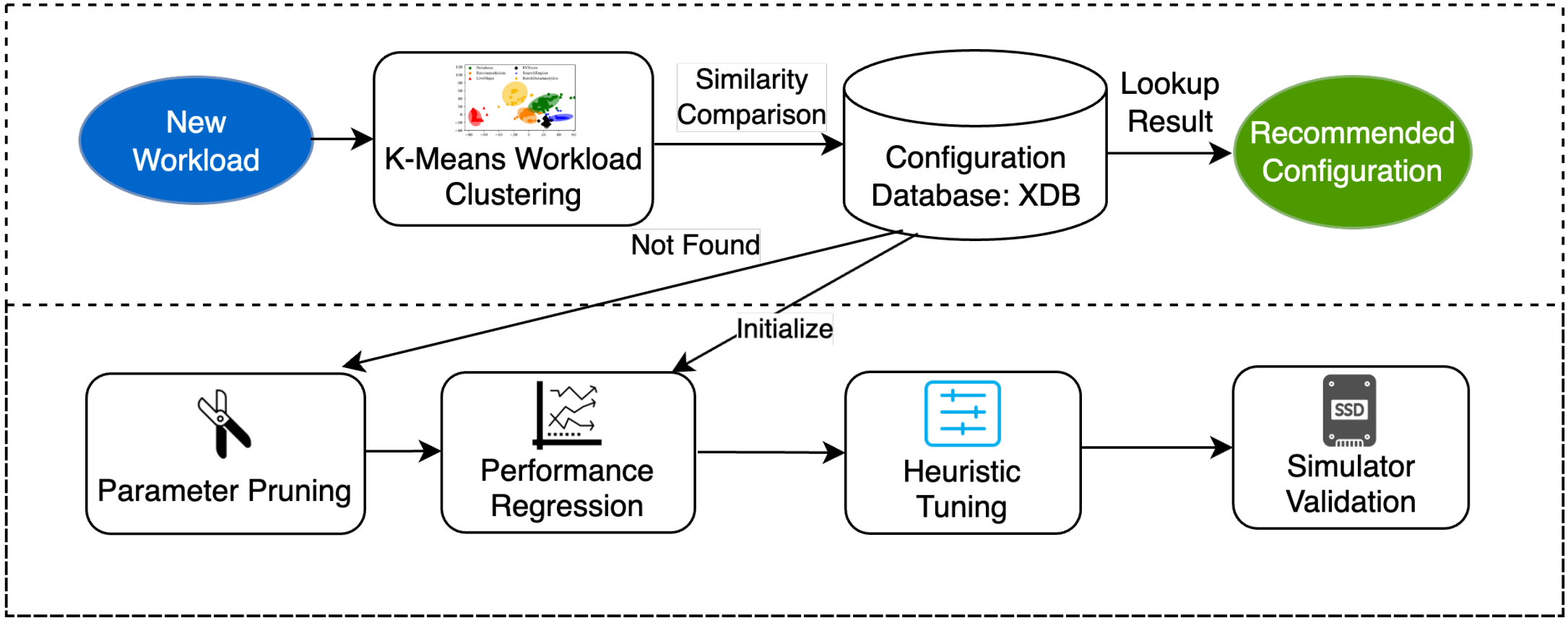
The Workflow of Learning Optimized SSD Configurations



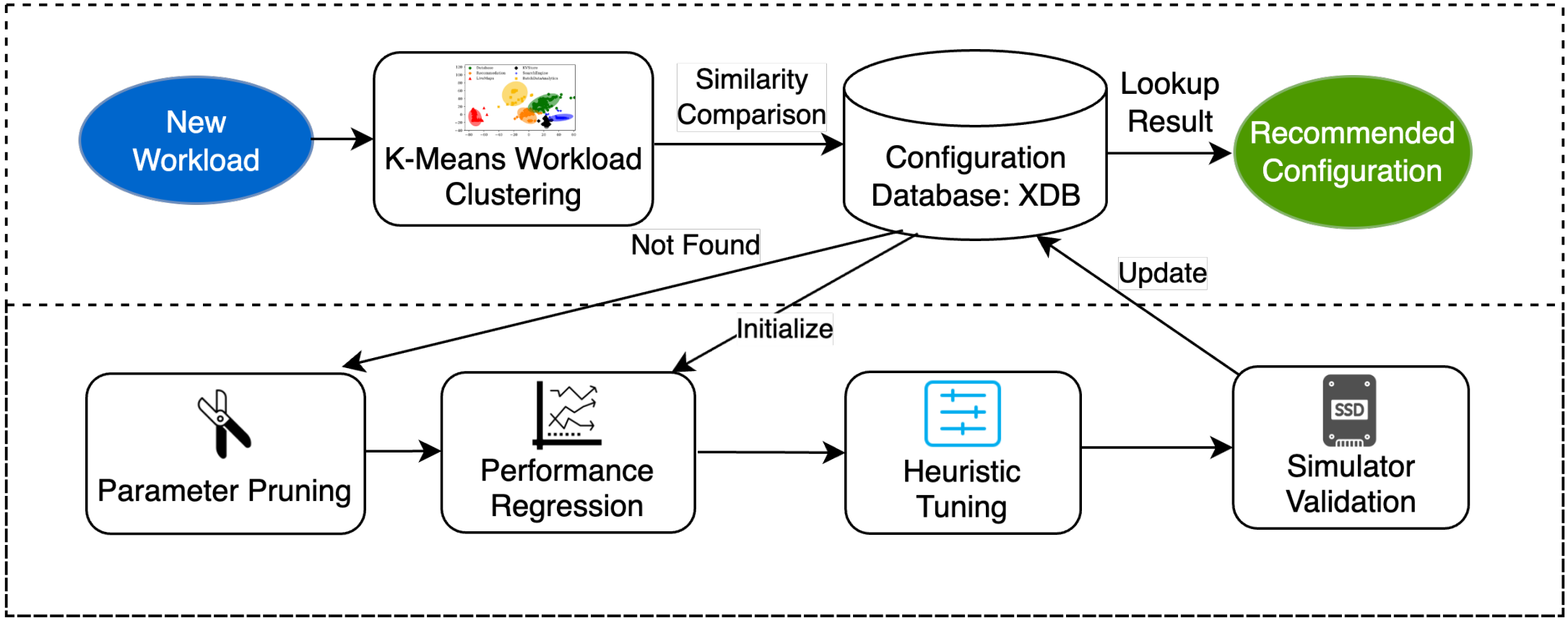
The Workflow of Learning Optimized SSD Configurations



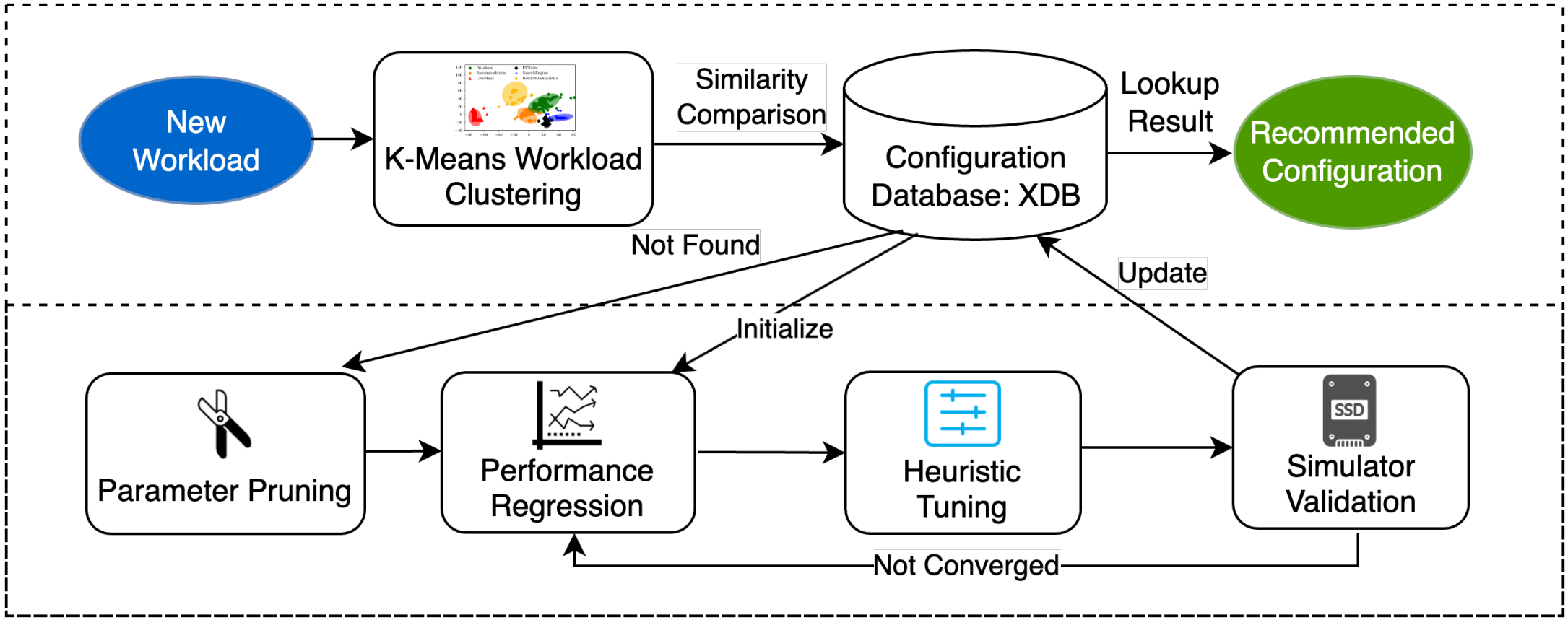
The Workflow of Learning Optimized SSD Configurations



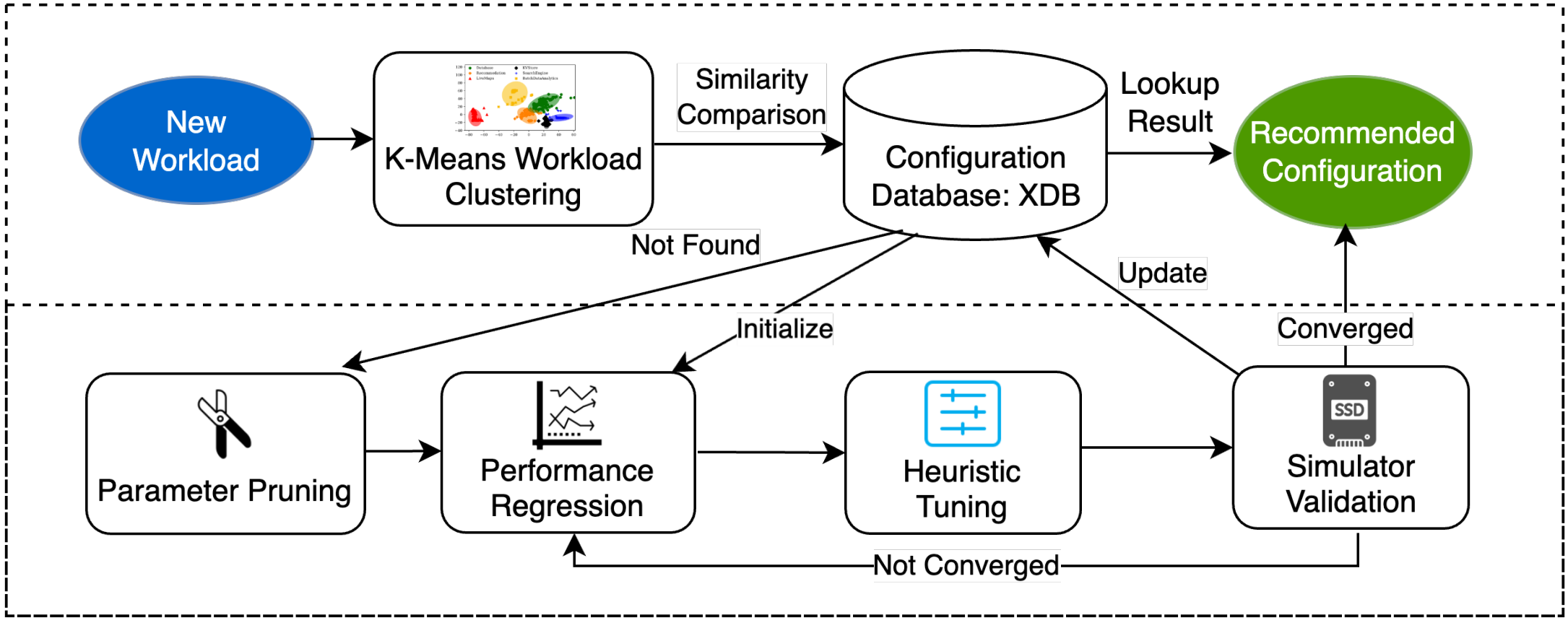
The Workflow of Learning Optimized SSD Configurations



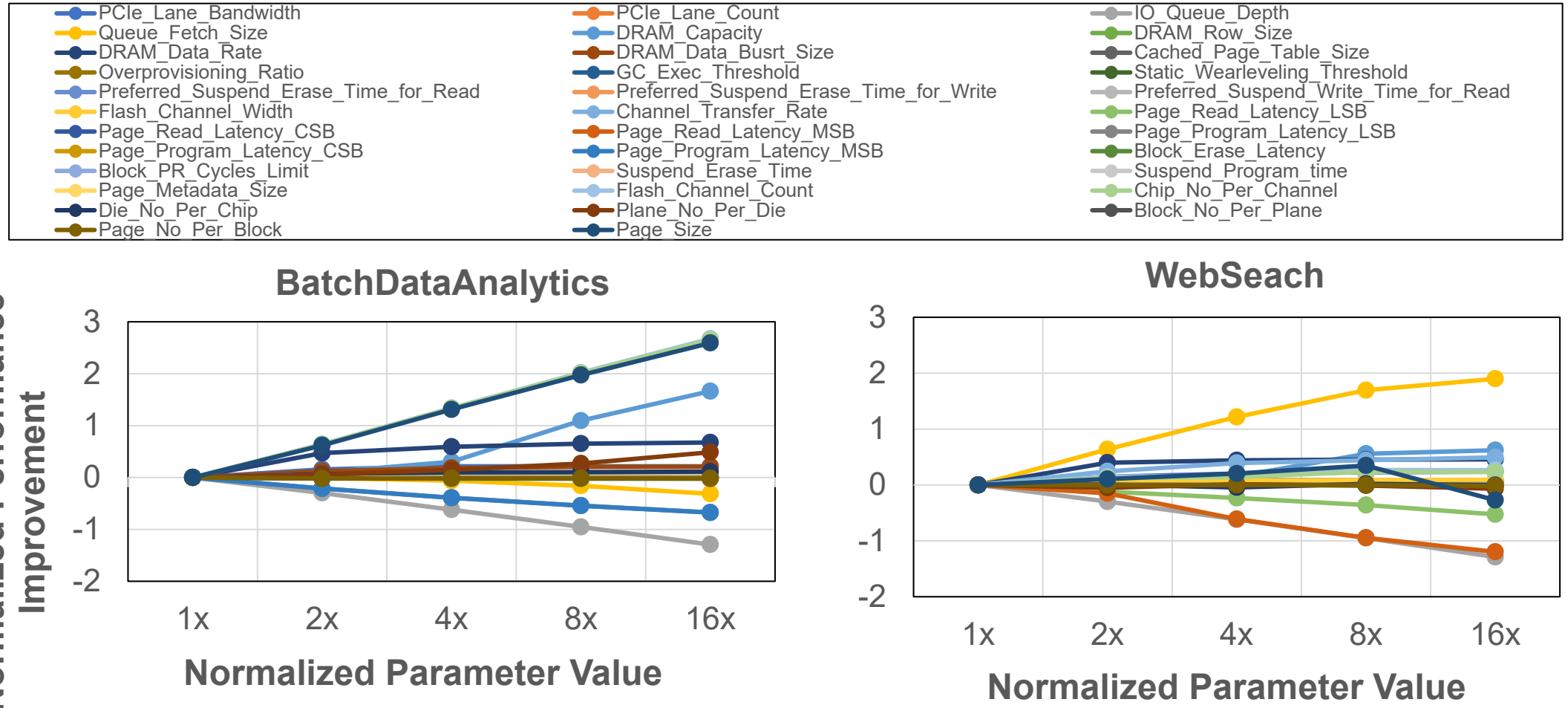
The Workflow of Learning Optimized SSD Configurations



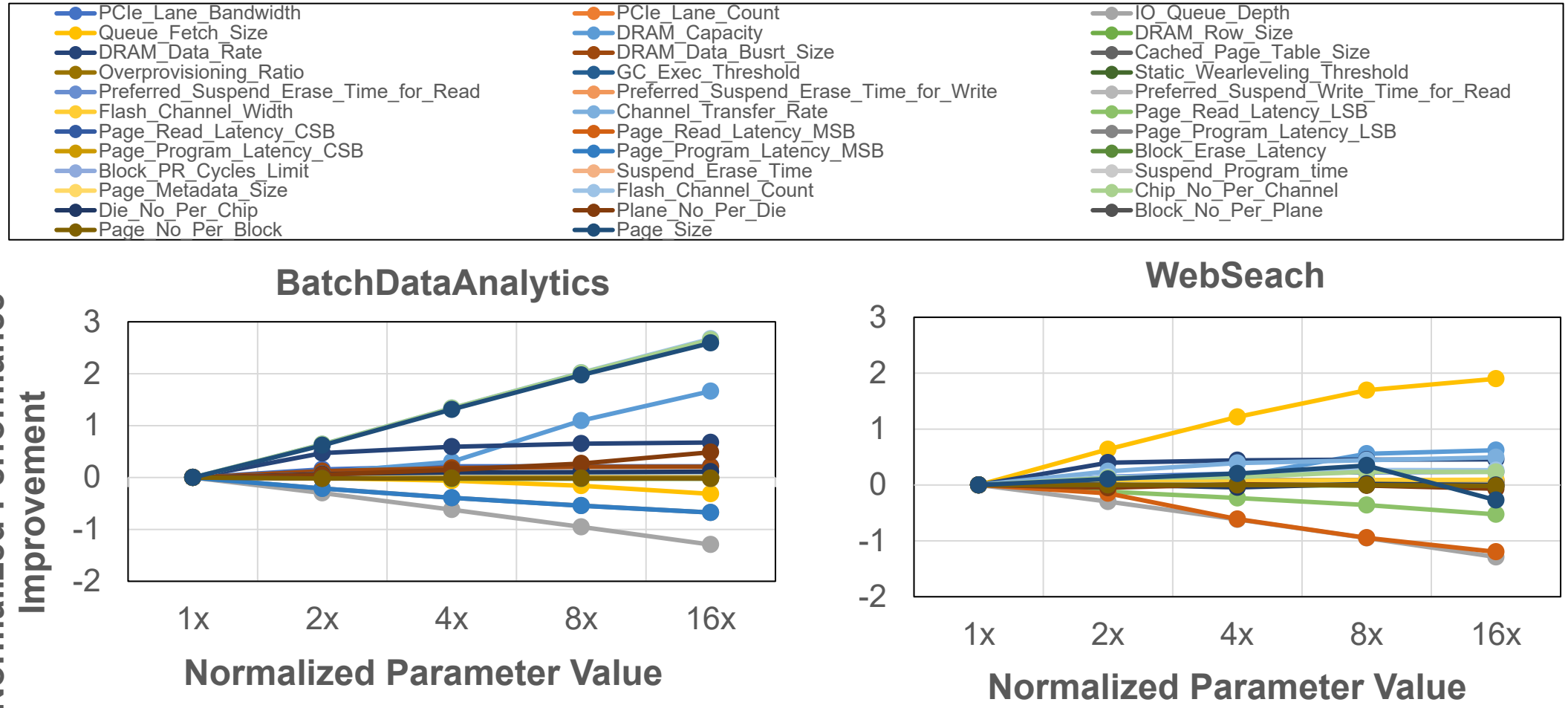
The Workflow of Learning Optimized SSD Configurations



Parameter Pruning for Reduced Search Space and Training Overhead



Parameter Pruning for Reduced Search Space and Training Overhead



We remove 45% of the parameters with parameter pruning

Performance Benefit of Learned SSD Configurations

Target Workload

Test Workload: Normalized Performance	Recomm	BatchAnalytics	WebSearch	CloudStorage	LiveMaps
Recomm	**1.25/1.26**	1.19/1.21	0.99/1.04	1.10/1.13	1.18/1.17
BatchAnalytics	1.14/1.13	**1.26/1.27**	1.14/1.15	1.08/1.08	1.12/1.11
WebSearch	0.87/0.88	0.75/0.76	**1.61/1.61**	1.01/1.00	1.00/1.00
CloudStorage	1.34/1.36	1.21/1.22	1.04/1.05	**1.48/1.46**	1.15/1.14
LiveMaps	1.16/1.16	1.04/1.06	0.97/0.98	1.07/1.07	**1.24/1.23**

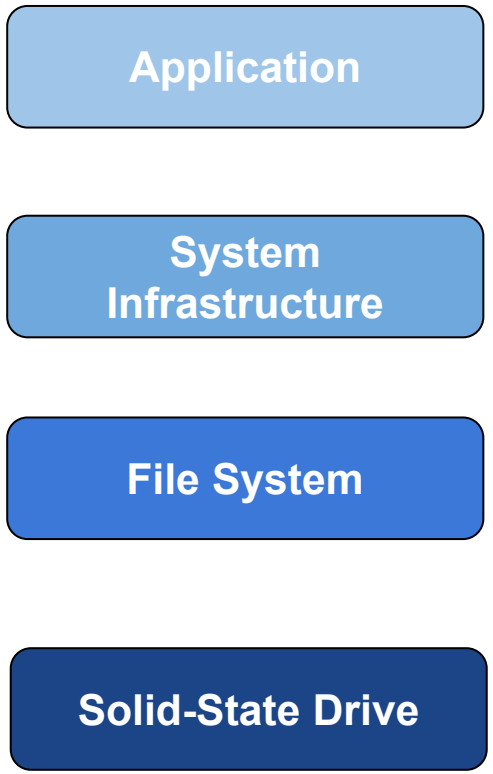
Performance Benefit of Learned SSD Configurations

Target Workload

Test Workload / Normalized Performance	Recomm	BatchAnalytics	WebSearch	CloudStorage	LiveMaps
Recomm	**1.25/1.26**	1.19/1.21	0.99/1.04	1.10/1.13	1.18/1.17
BatchAnalytics	1.14/1.13	**1.26/1.27**	1.14/1.15	1.08/1.08	1.12/1.11
WebSearch	0.87/0.88	0.75/0.76	**1.61/1.61**	1.01/1.00	1.00/1.00
CloudStorage	1.34/1.36	1.21/1.22	1.04/1.05	**1.48/1.46**	1.15/1.14
LiveMaps	1.16/1.16	1.04/1.06	0.97/0.98	1.07/1.07	**1.24/1.23**

We improve 1.37x for target workloads and 1.08x for non-target workloads

Building A Learning-based Storage Ecosystem



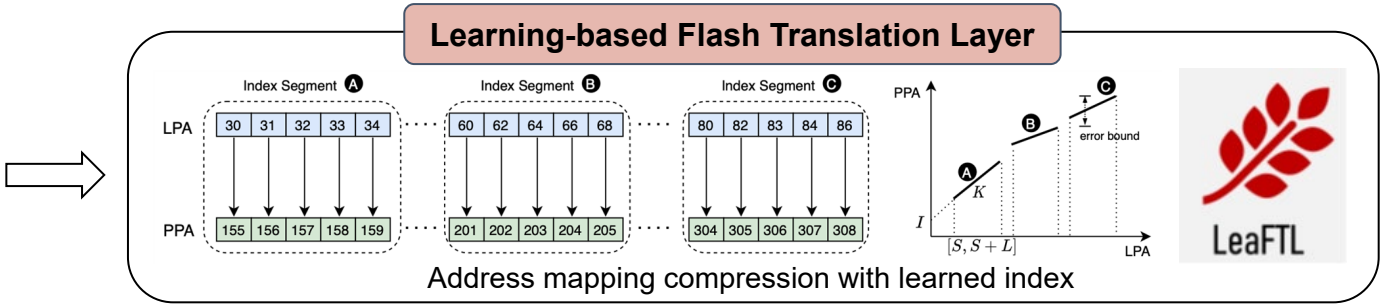
Building A Learning-based Storage Ecosystem

Application

System Infrastructure

File System

Solid-State Drive



LeaFTL (ASPLOS ‘23)

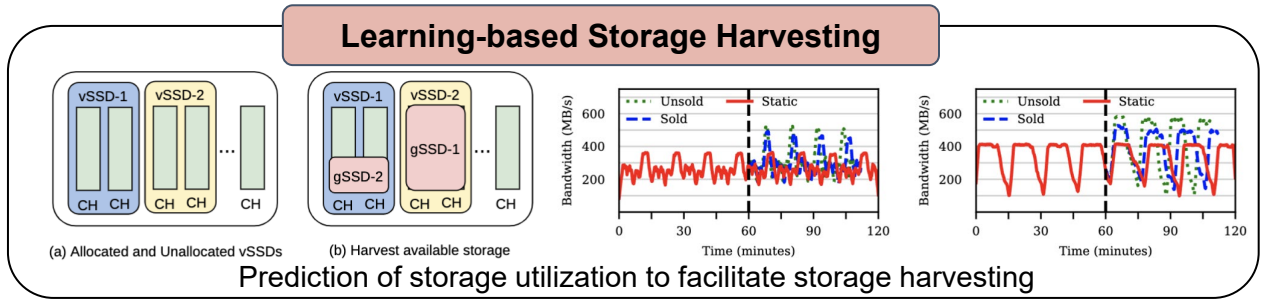
Building A Learning-based Storage Ecosystem

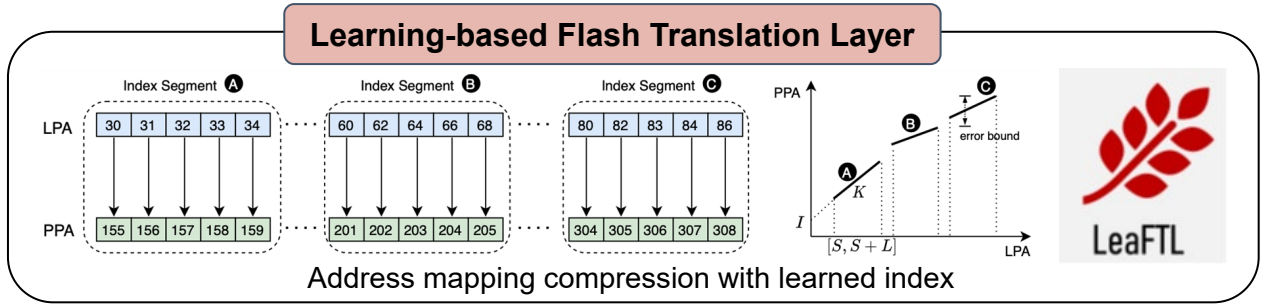
Application

System Infrastructure

File System

Solid-State Drive





BlockFlex (OSDI '22)

LeaFTL (ASPLOS '23)

Building A Learning-based Storage Ecosystem

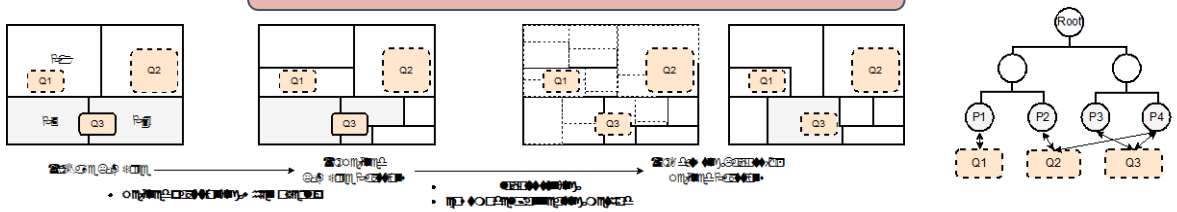
Application

System Infrastructure

File System

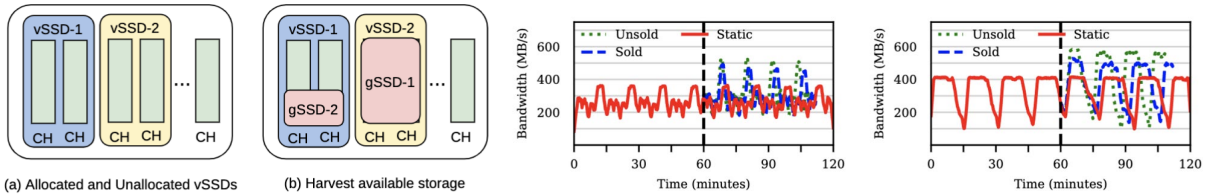
Solid-State Drive

Learning-based Spatial Indexing



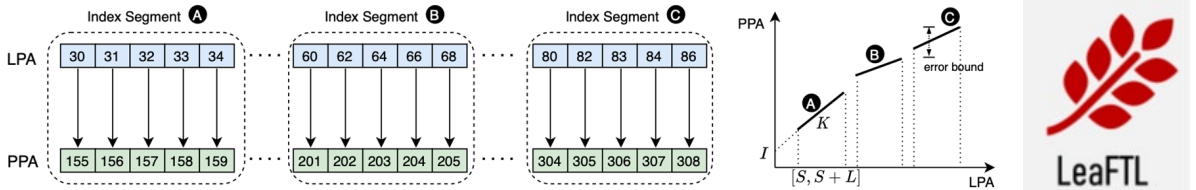
Query-aware methods to optimize data partitioning for spatial datasets

Learning-based Storage Harvesting



Prediction of storage utilization to facilitate storage harvesting

Learning-based Flash Translation Layer



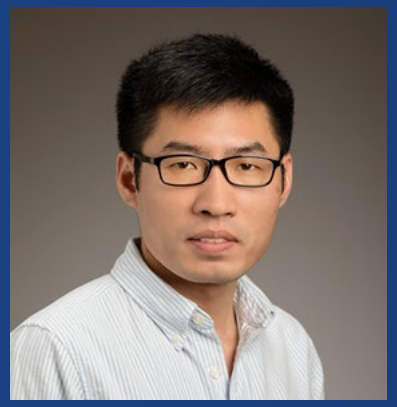
Address mapping compression with learned index

Learning-based Spatial Indexing (Coming soon)

BlockFlex (OSDI '22)

LeaFTL (ASPLOS '23)

Thank You!



Jian Huang **Daixuan Li** **Jinghan Sun**

Systems Platform Research Group

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN