

Towards Learning-based Storage Systems: A Holistic Approach

Jinghan Sun Daixuan Li Jian Huang

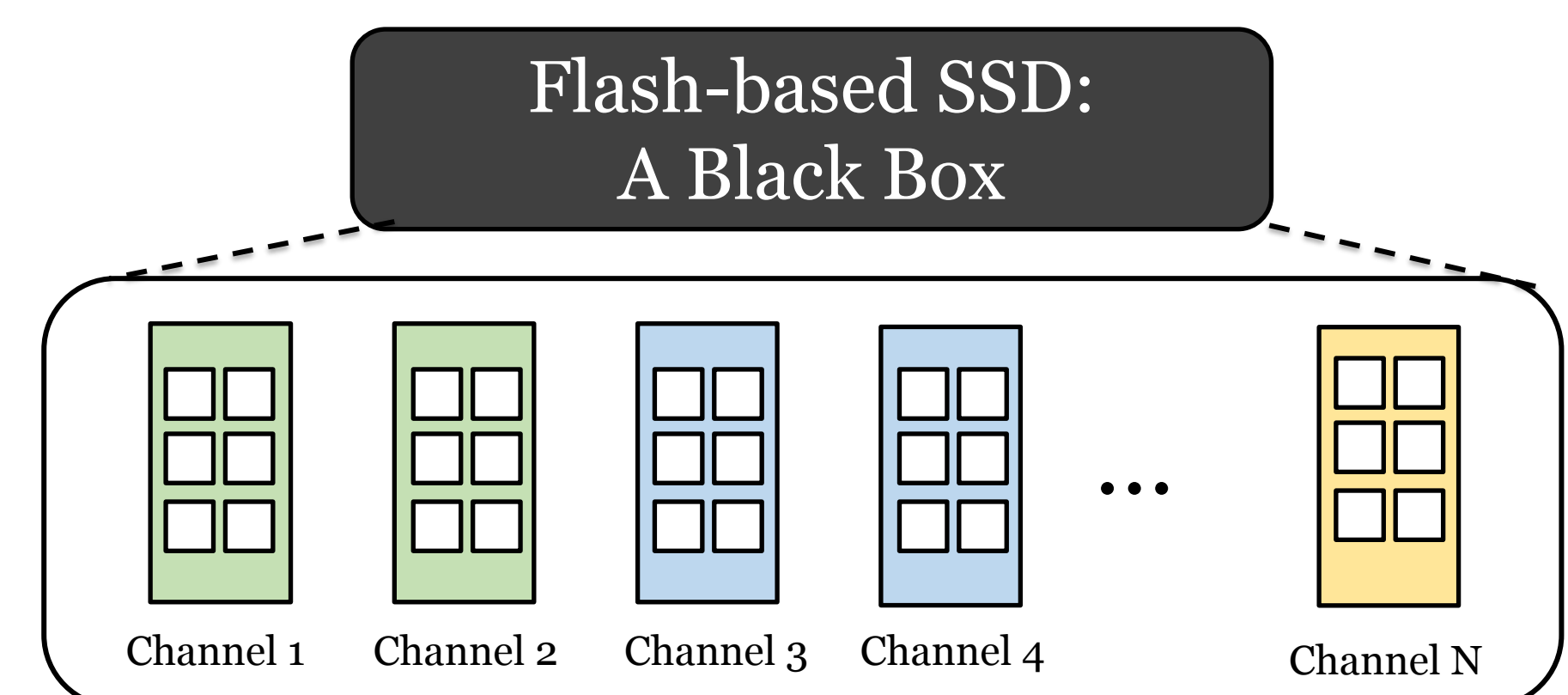
University of Illinois at Urbana-Champaign



Workshop on ML for Systems

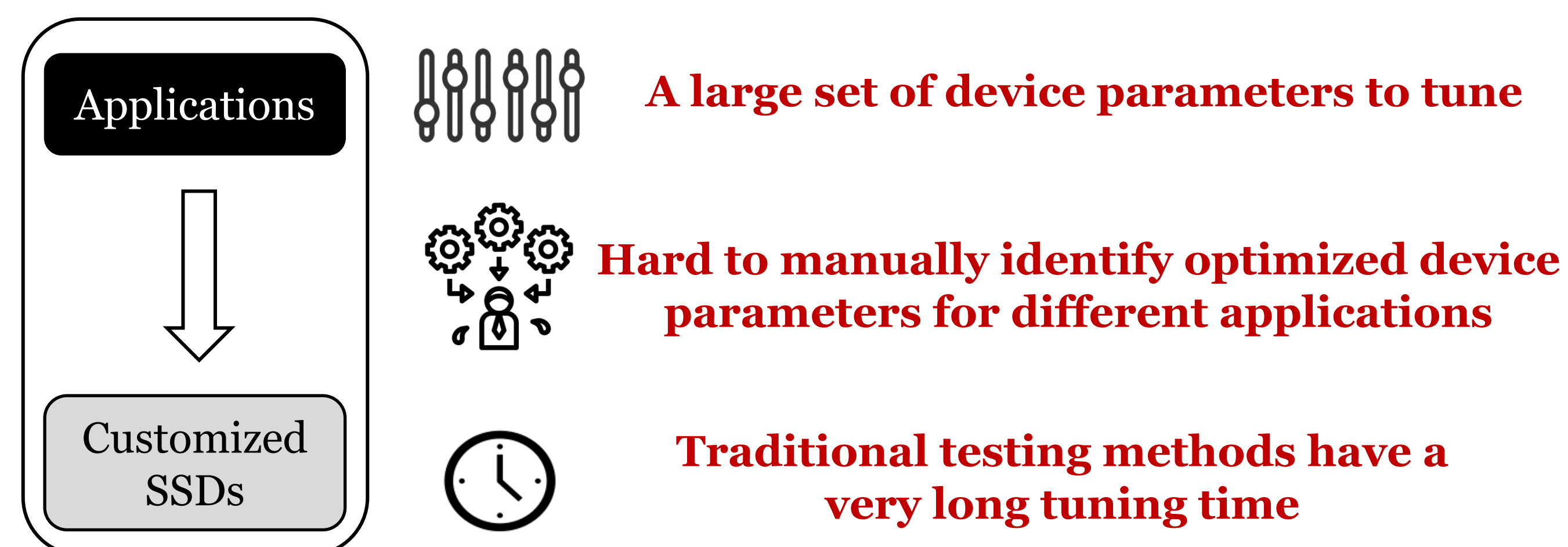
LEARNING TO DRIVE SOFTWARE-DEFINED STORAGE

1. Software-Defined Storage

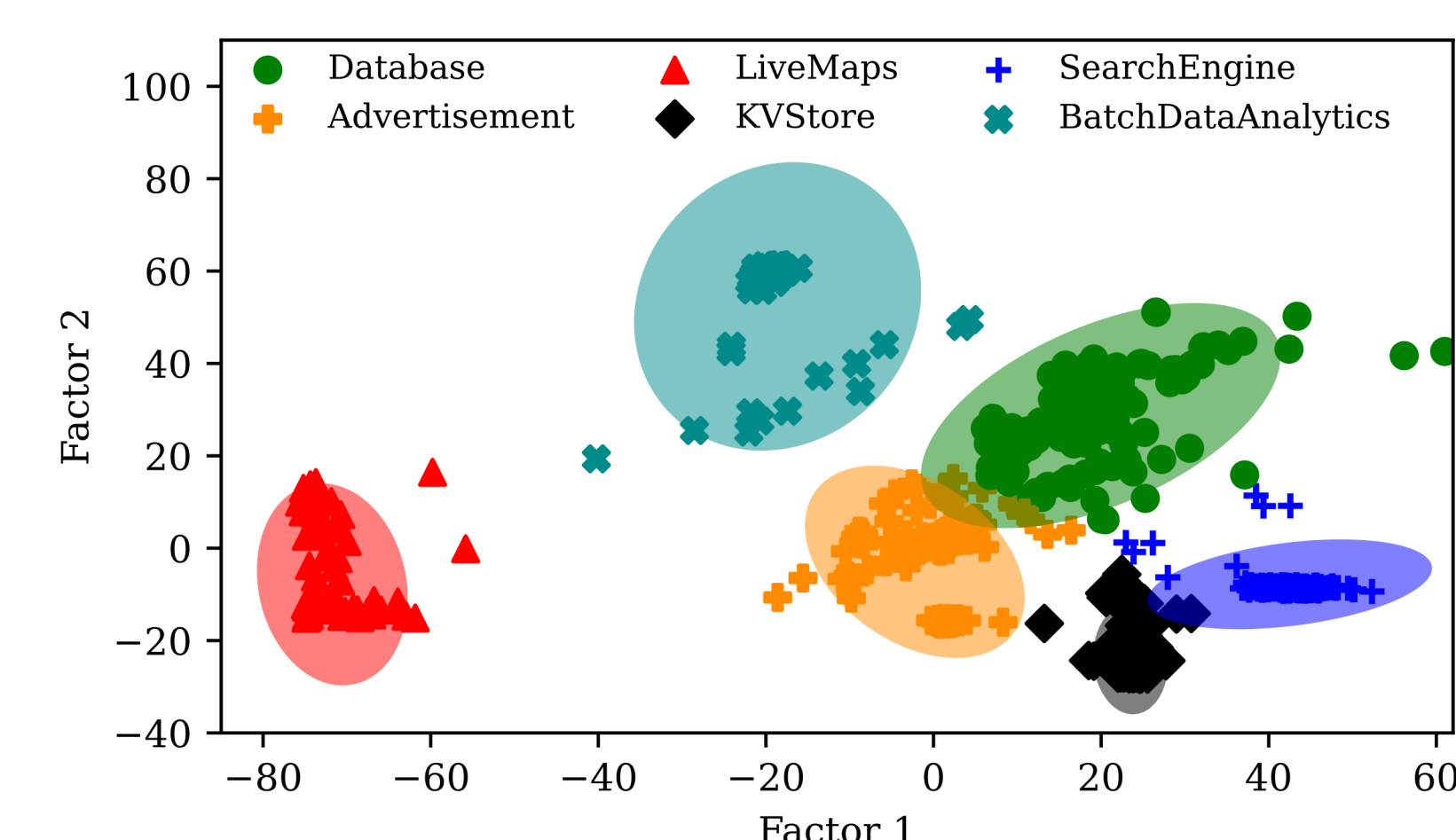


Software-Defined Storage Opens the Blackbox of SSD

2. Challenges of Building Software-Defined Storage

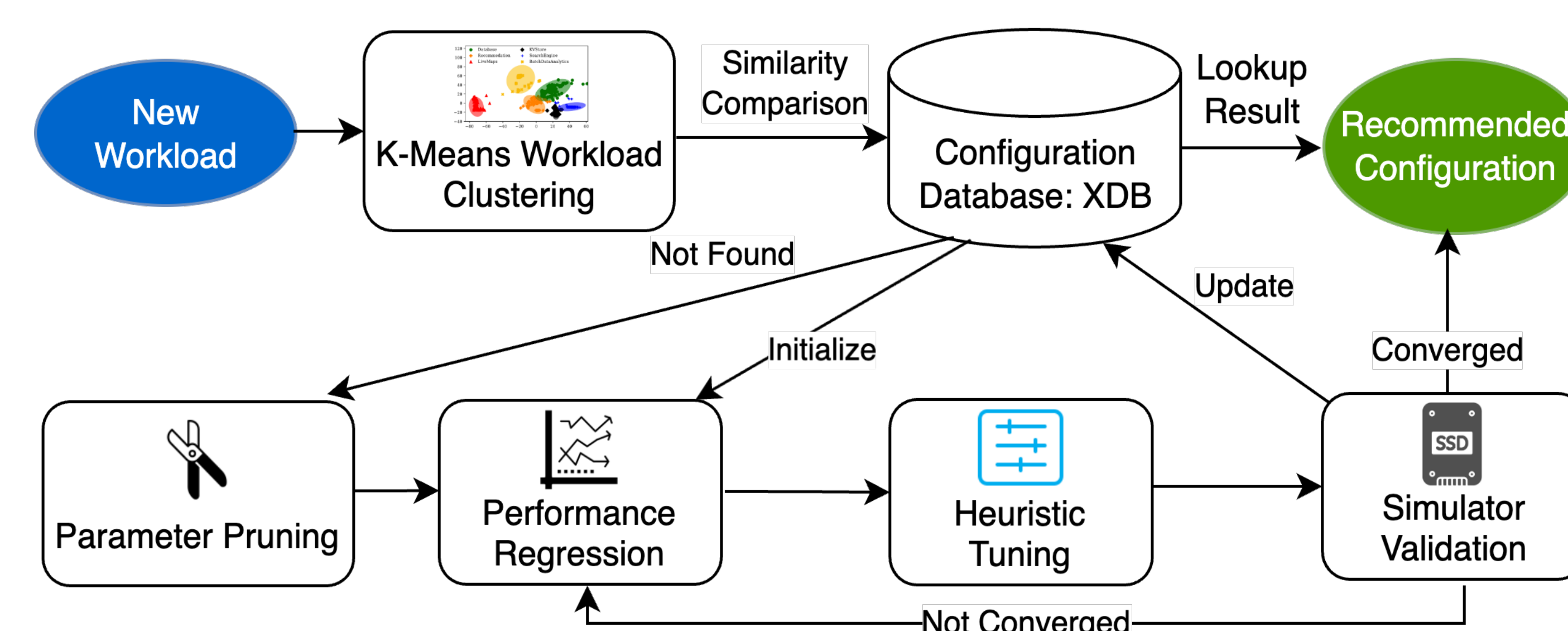


3. Application Characterization with Clustering



The Learning-based Approach is Feasible to Study Storage Workloads

4. Workflow of Our Learning Framework



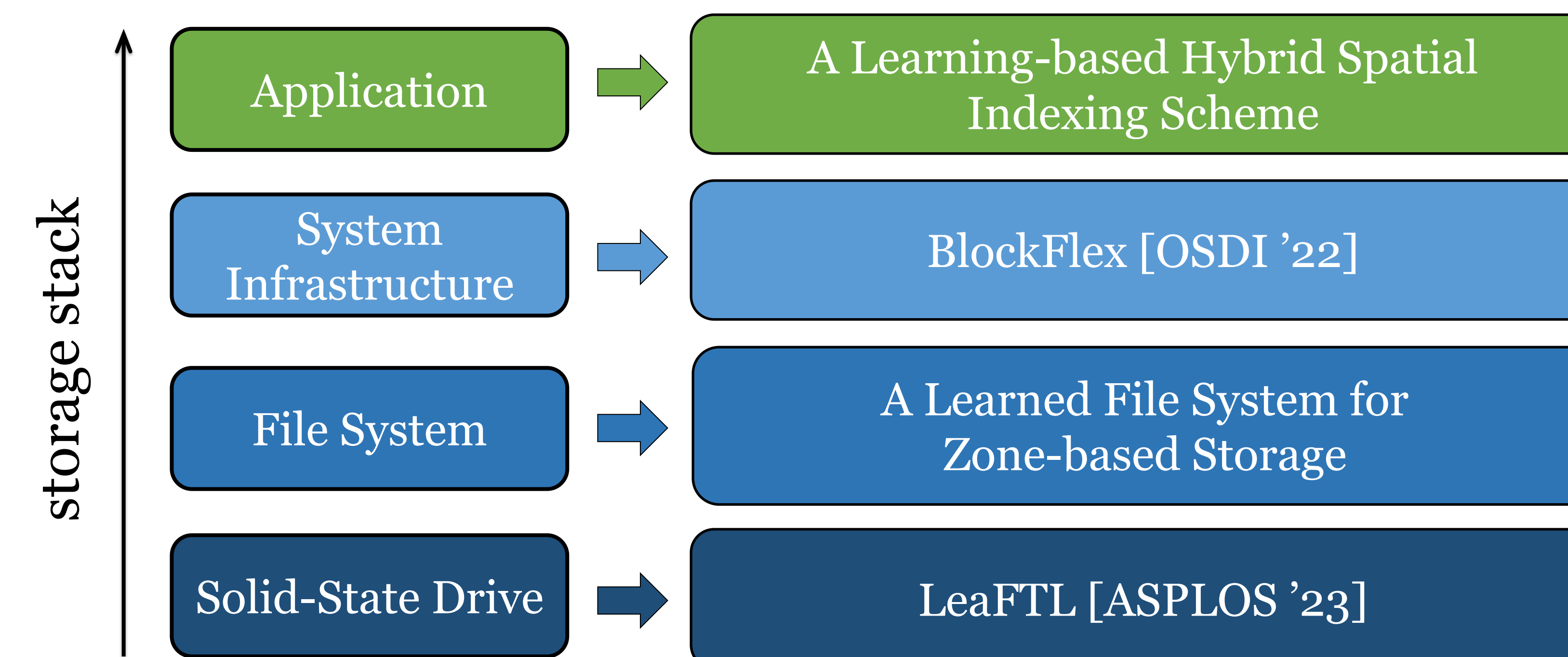
5. Performance Improvement of the Learned SSD Configurations

Target Workloads

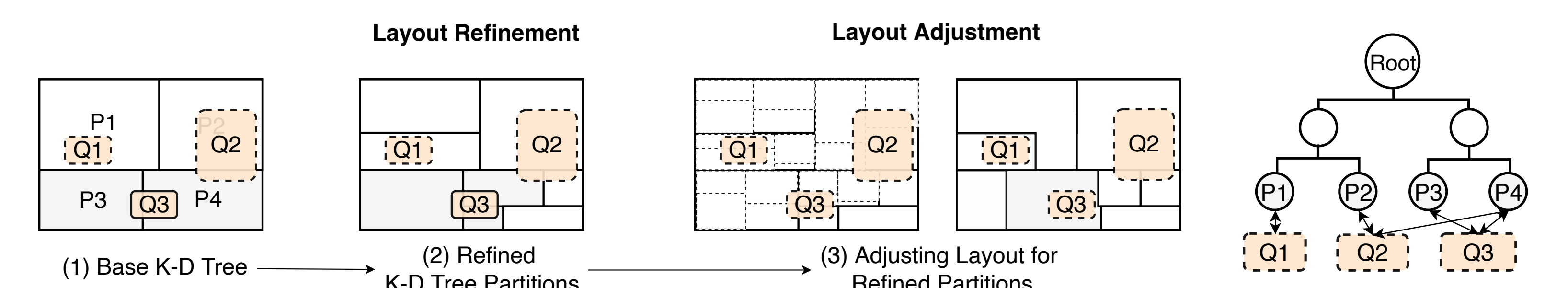
Performance Improvement	Recomm	BatchAnalytics	WebSearch	CloudStorage	LiveMaps
Recomm	1.25/1.26	1.19/1.21	0.99/1.04	1.10/1.13	1.18/1.17
BatchAnalytics	1.14/1.13	1.26/1.27	1.14/1.15	1.08/1.08	1.12/1.11
WebSearch	0.87/0.88	0.75/0.76	1.61/1.61	1.01/1.00	1.00/1.00
CloudStorage	1.34/1.36	1.21/1.22	1.04/1.05	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 Obtain **1.37x** Performance Improvement for Target Workloads and **1.08x** for Non-Target Workloads

A LEARNING-BASED STORAGE ECOSYSTEM

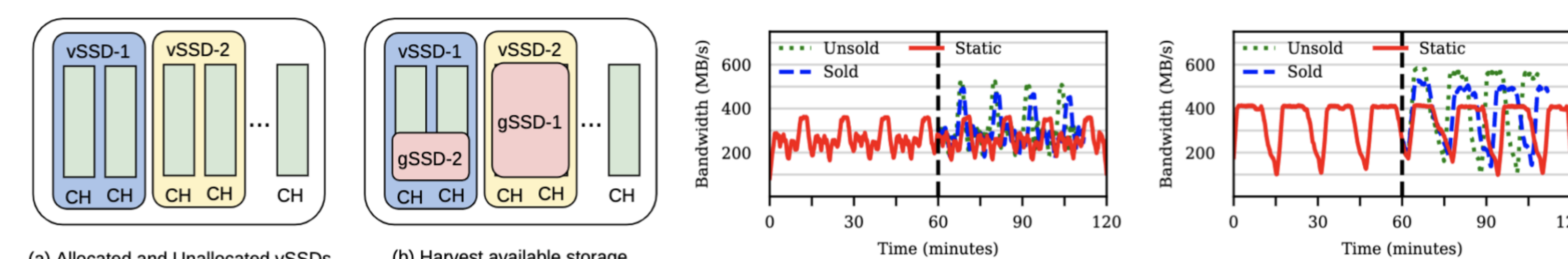


Hylinx: Learning-based Spatial Indexing



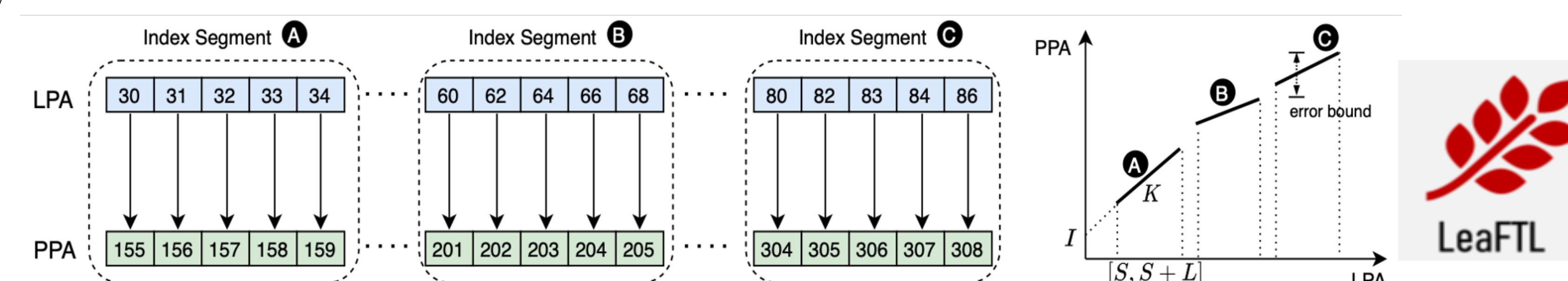
A Query-aware Approach to Optimize Data Partitioning for Spatial Datasets

BlockFlex: Learning-based Storage Harvesting



Predicting Storage Utilization and Demand to Facilitate Storage Harvesting

LeaFTL: A Learning-based Flash Translation Layer



Reducing the Memory Footprint of SSD Mapping Table with Learned Indexes

Test Workloads