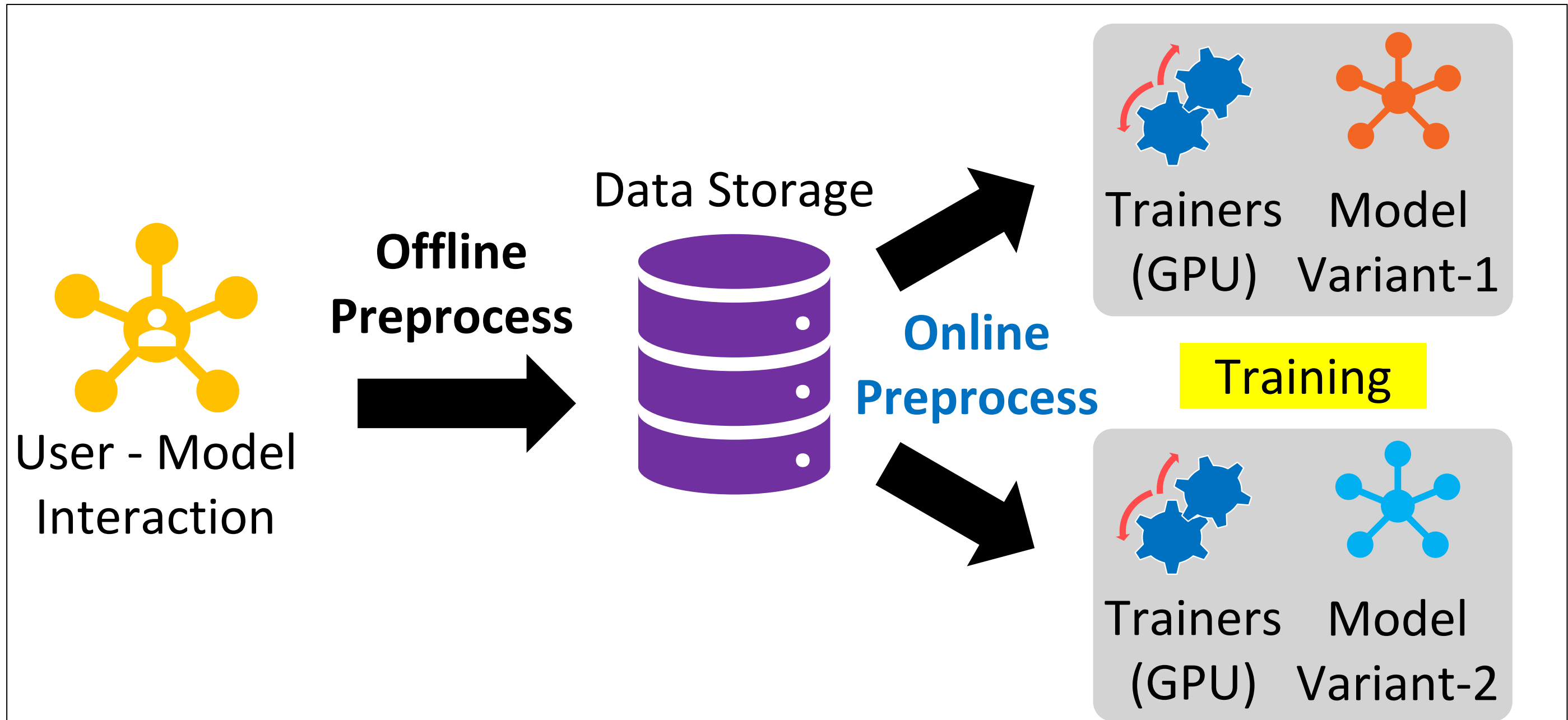


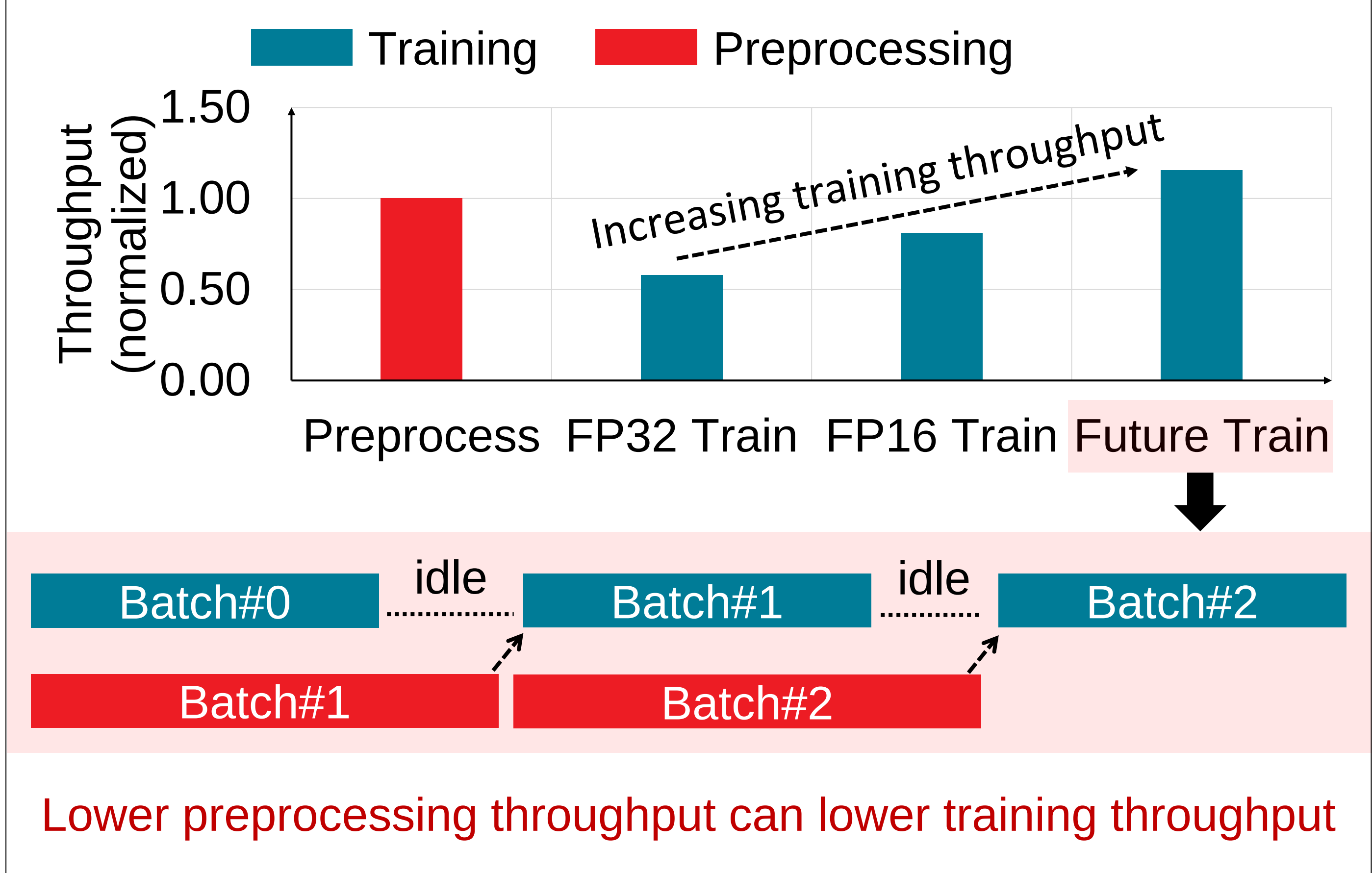
Accelerating ML Preprocessing via Processing-in-Memory

Mohamed Assem Ibrahim and Shaizeen Aga
Advanced Micro Devices, Inc.

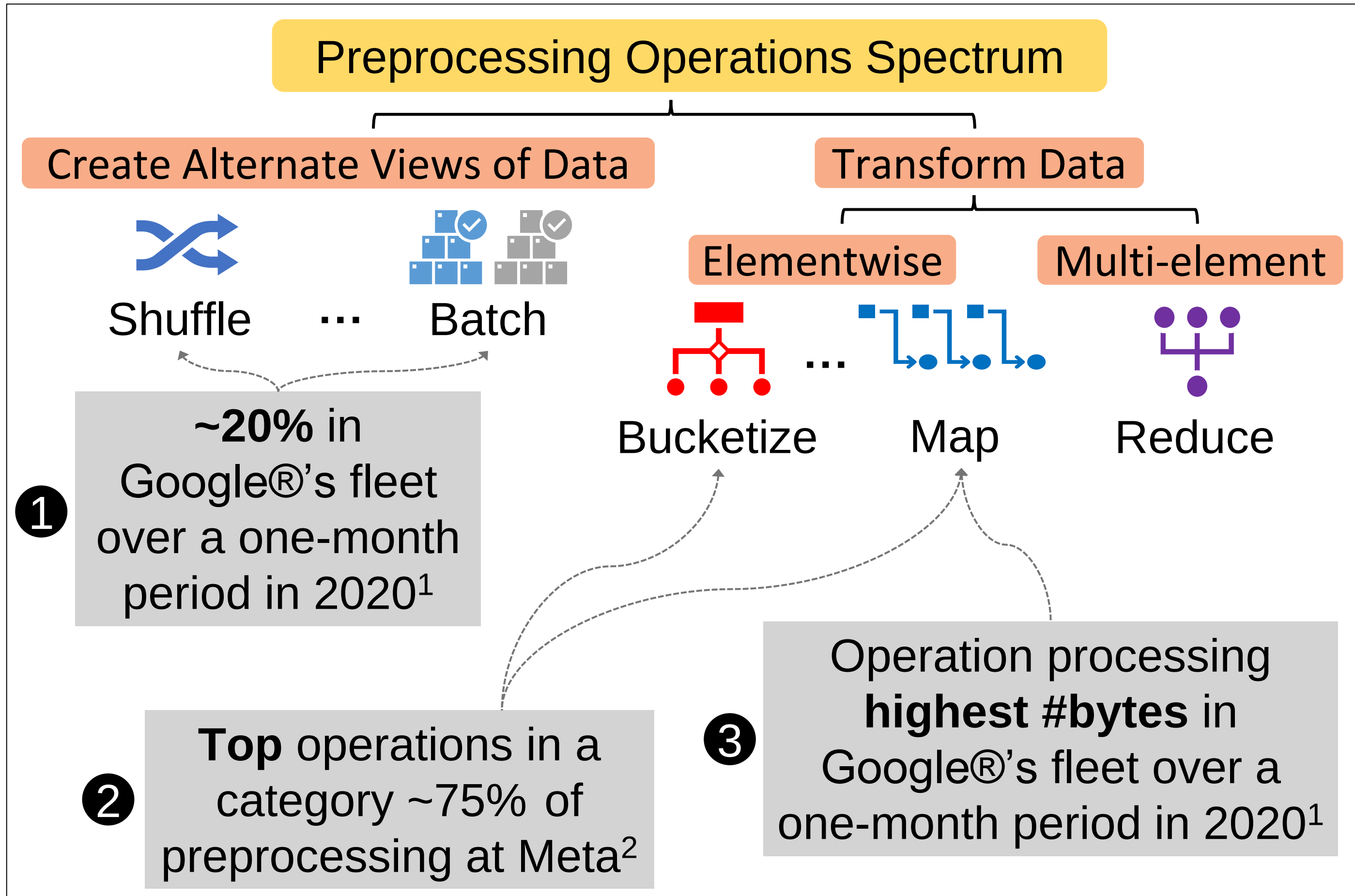
ML Preprocessing = Third Pillar of ML



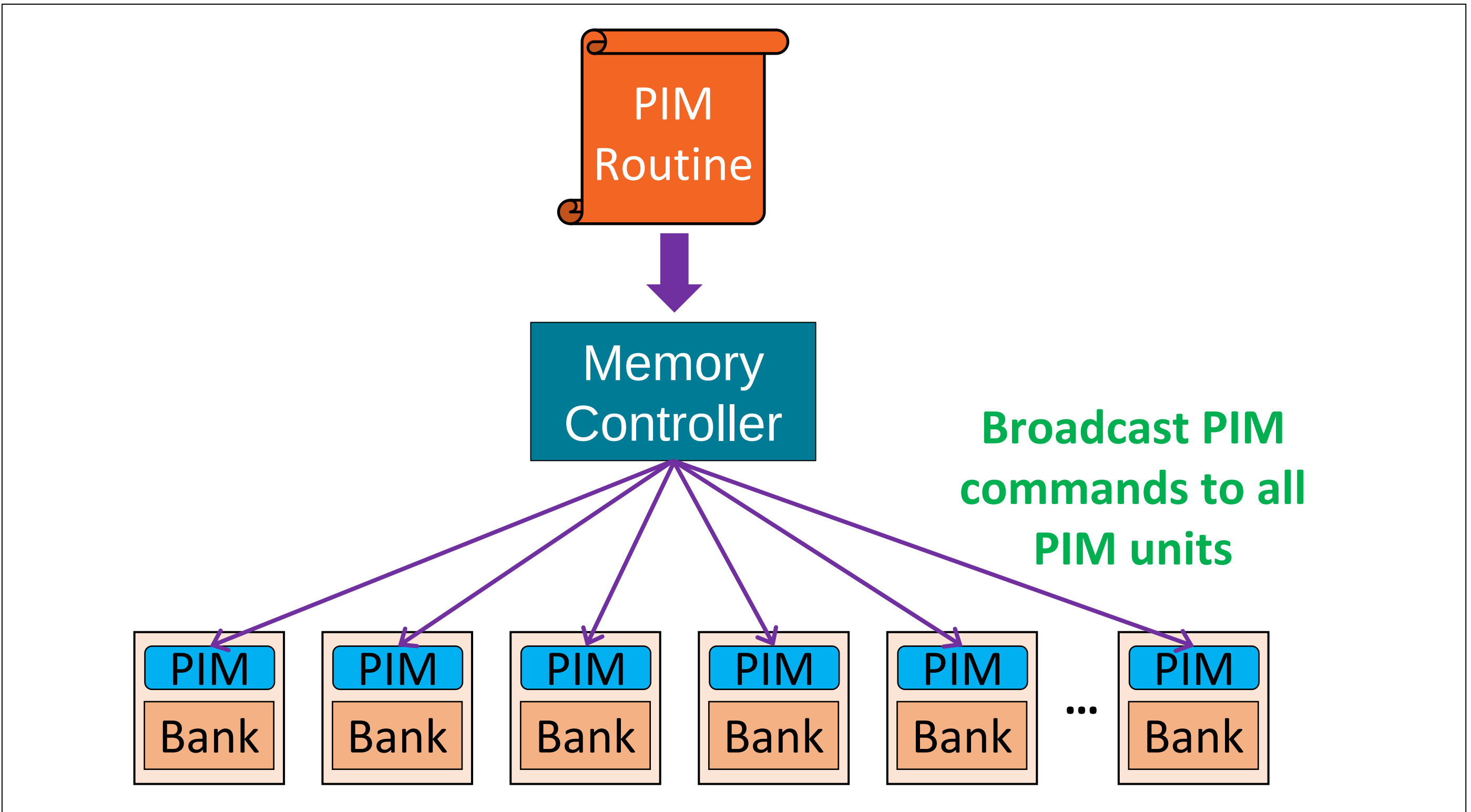
- Focus: **Online preprocessing**
- Preprocessing consumes considerable ML resources
- Large data volumes consumed by preprocessing



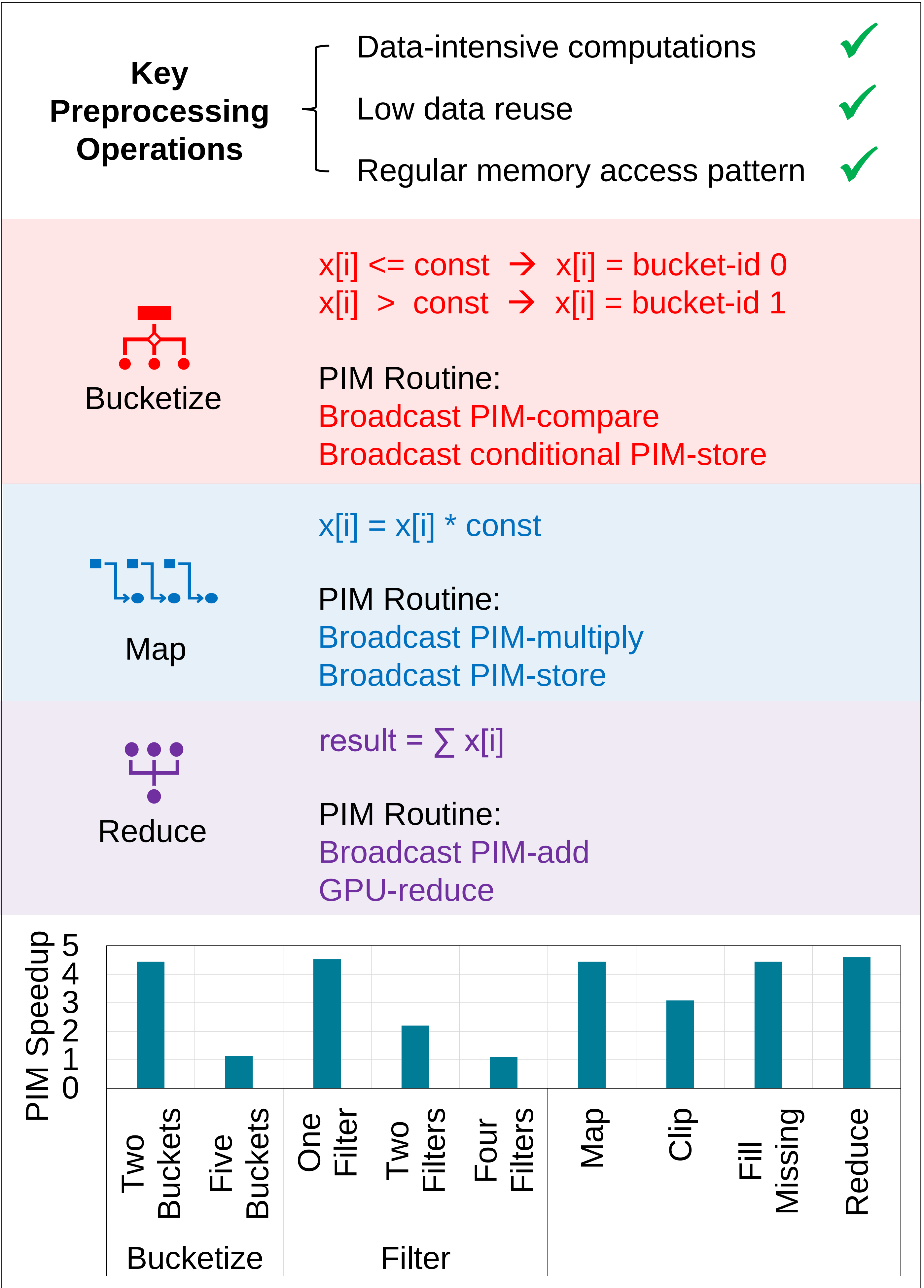
ML Preprocessing is Data-intensive



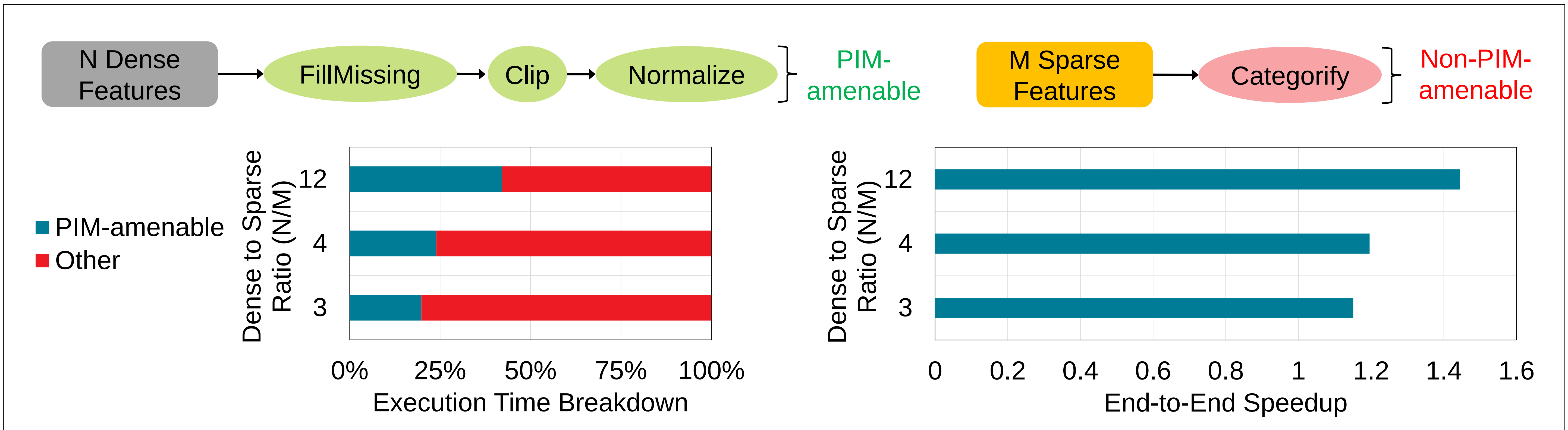
PIM Overview



PIM for ML Preprocessing



Evaluating Criteo 1TB Click Logs Preprocessing Pipeline³



[1] tf.data: A Machine Learning Data Processing Framework, Murray et. al.
[2] Understanding Data Storage and Ingestion for Large-Scale Deep Recommendation Model Training, Zhao et. al.
[3] <https://ailab.criteo.com/download-criteo-1tb-click-logs-dataset/>
[4] Hardware Architecture and Software Stack for PIM Based on Commercial DRAM Technology, Lee et. al.