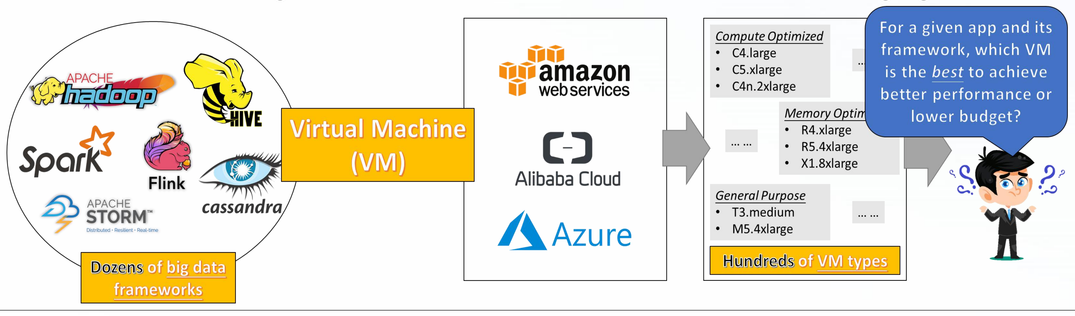
Best VM Selection for Big Data Applications across Multiple Frameworks by Transfer Learning

Yuewen Wu, Heng Wu, Yuanjia Xu, Yi Hu, Wenbo Zhang,
Hua Zhong, Tao Huang

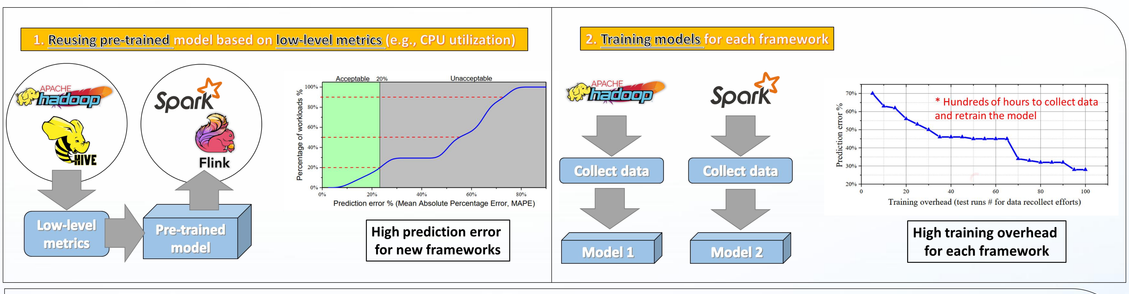
In Proceeding of the 50th International Conference on Parallel Processing, ICPP 2021.

Contact: Wu Yuewen, 18600612053, wuyuewen@otcaix.iscas.ac.cn

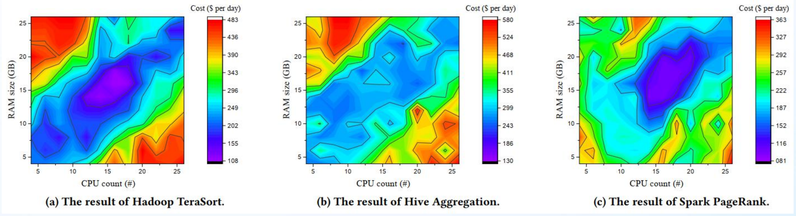
Selecting the best VM for multiple frameworks is challenging



Limitations of existing machine learning approaches



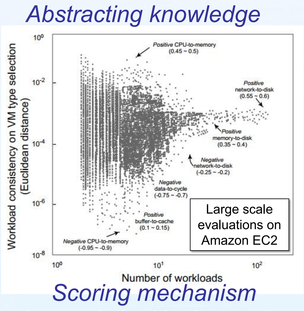
The core finding: *knowledge* across multiple frameworks



Low-level metrics have **high-level similarities** (aka ***knowledge***) across frameworks
blue areas in heat maps show that multi-framework applications have similar CPU and RAM requirements

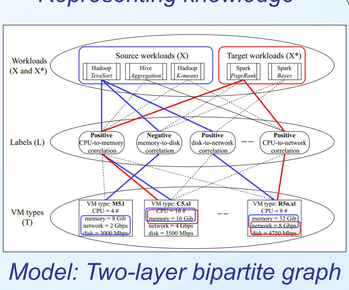
Vesta: reusing knowledge by transfer learning

Abstracting knowledge



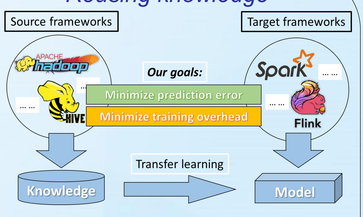
Scoring mechanism
X-axis: popularity Y-axis: consistency

Representing knowledge



Model: Two-layer bipartite graph
Blue: source Red: target

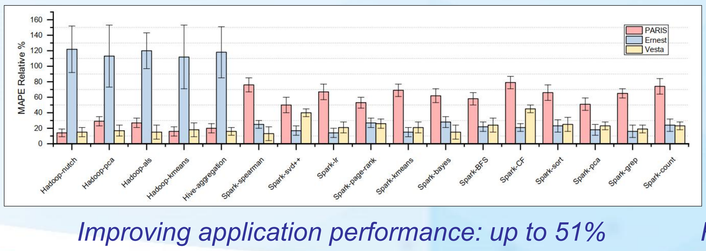
Reusing knowledge



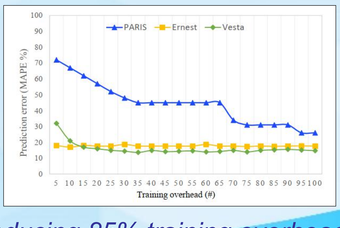
Solver: collective matrix factorization

$$\min_{U,F,U^*} \lambda \| U^* - U \|_F^2 + (1-\lambda) \| U^* - V \|_F^2 + R(U, V, U^*)$$

Improving application performance while reducing training overhead



Improving application performance: up to 51%



Reducing 85% training overhead