The 11th International Conference on Future Data and Security Engineering

FDSE 2024

VIETNAMESE-GERMAN UNIVERSITY, BEN CAT, BINH DUONG, VIETNAM

November 27-29, 2024

Springer · CCIS · VGU Vietnamese-German University · JKU LIT Secure and Correct Systems Lab · FAW · SPRINGER NATURE

FDSE Paper #22

JREP - A Job Runtime Ensemble Predictor for Improving Scheduling Performance on High Performance Computing Systems

Thanh Hoang Le Hai, Thin Nguyen Manh, Hung Nguyen Quang, and Nam Thoai

thanhhoang@hcmut.edu.vn

High Performance Computing Lab

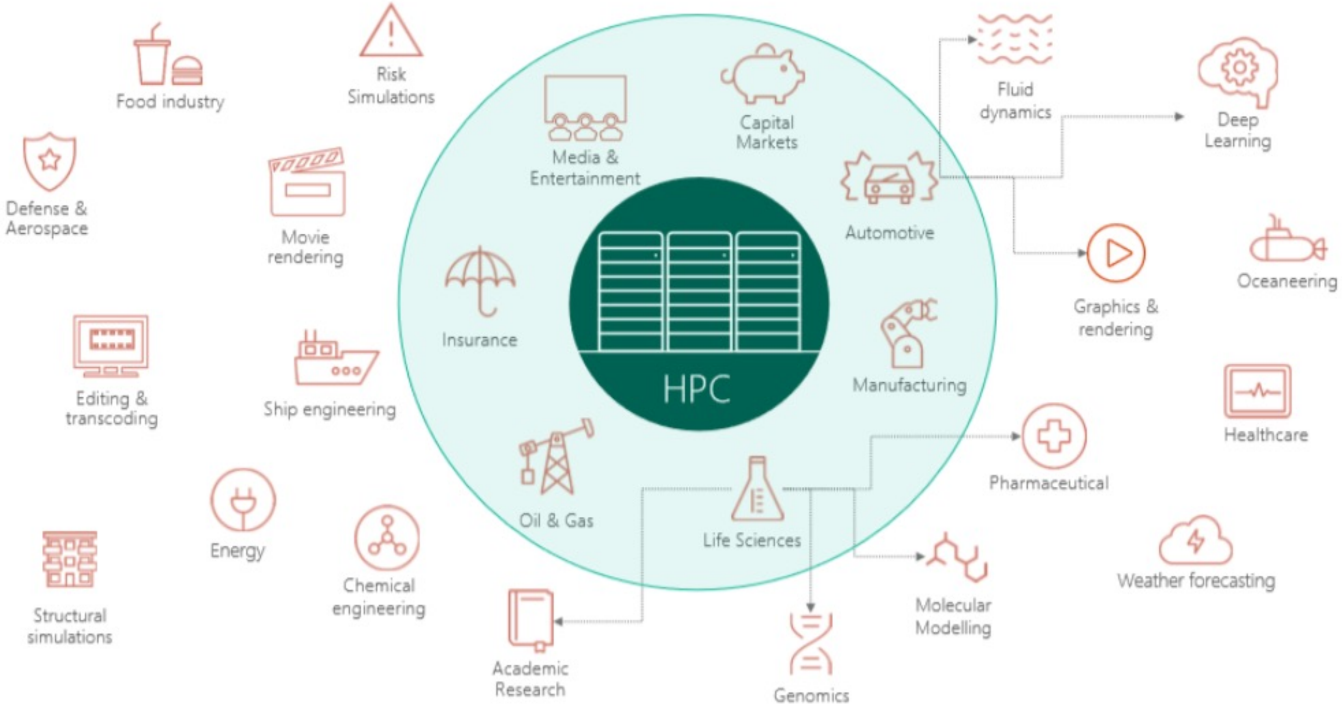
Faculty of Computer Science & Engineering, HCMUT, VNU-HCM

Today's Contents

1. Introduction

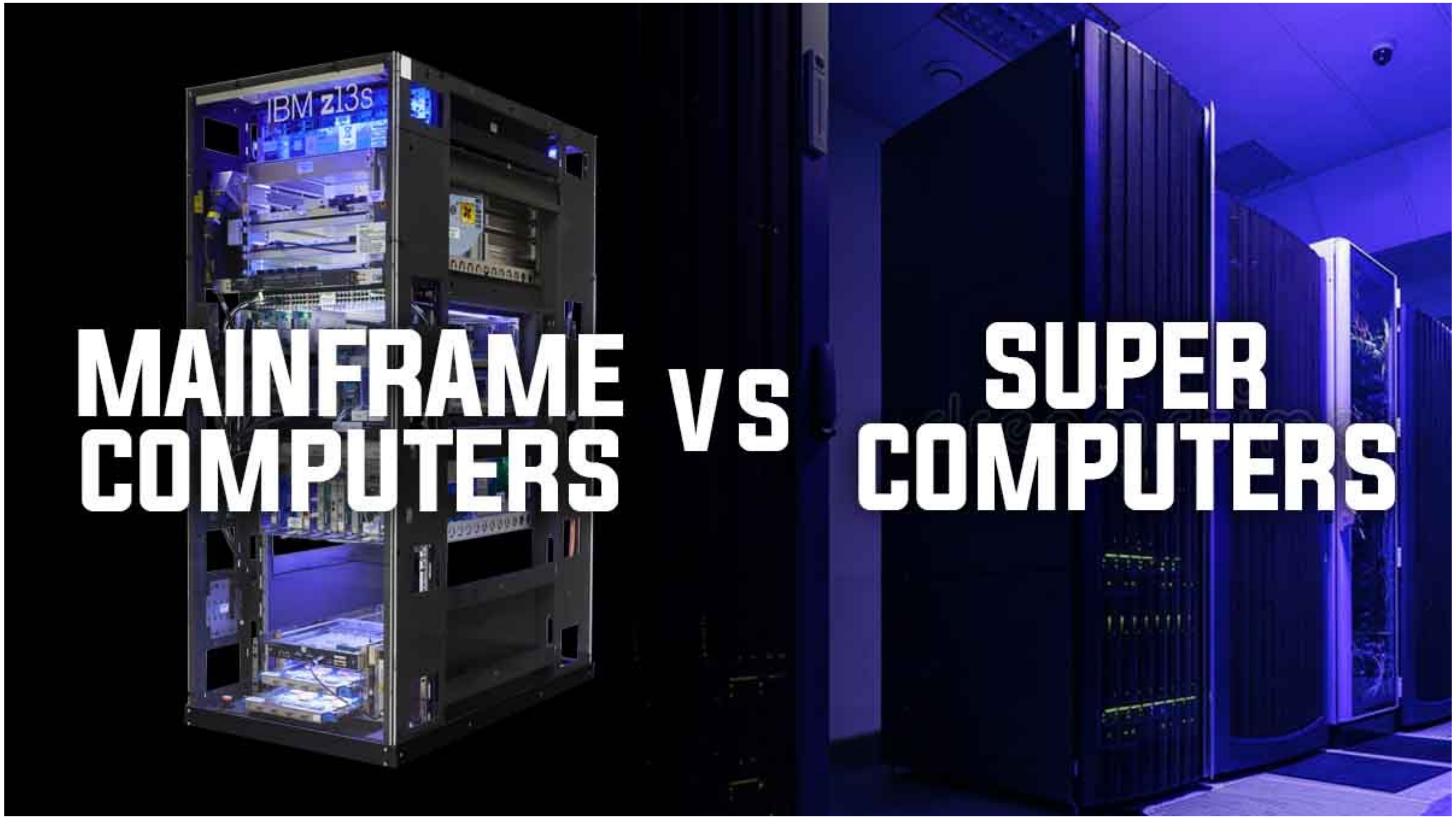
What is High Performance Computing?

HPC – High Performance Computing: Ability to process data and perform complex calculations at high speeds



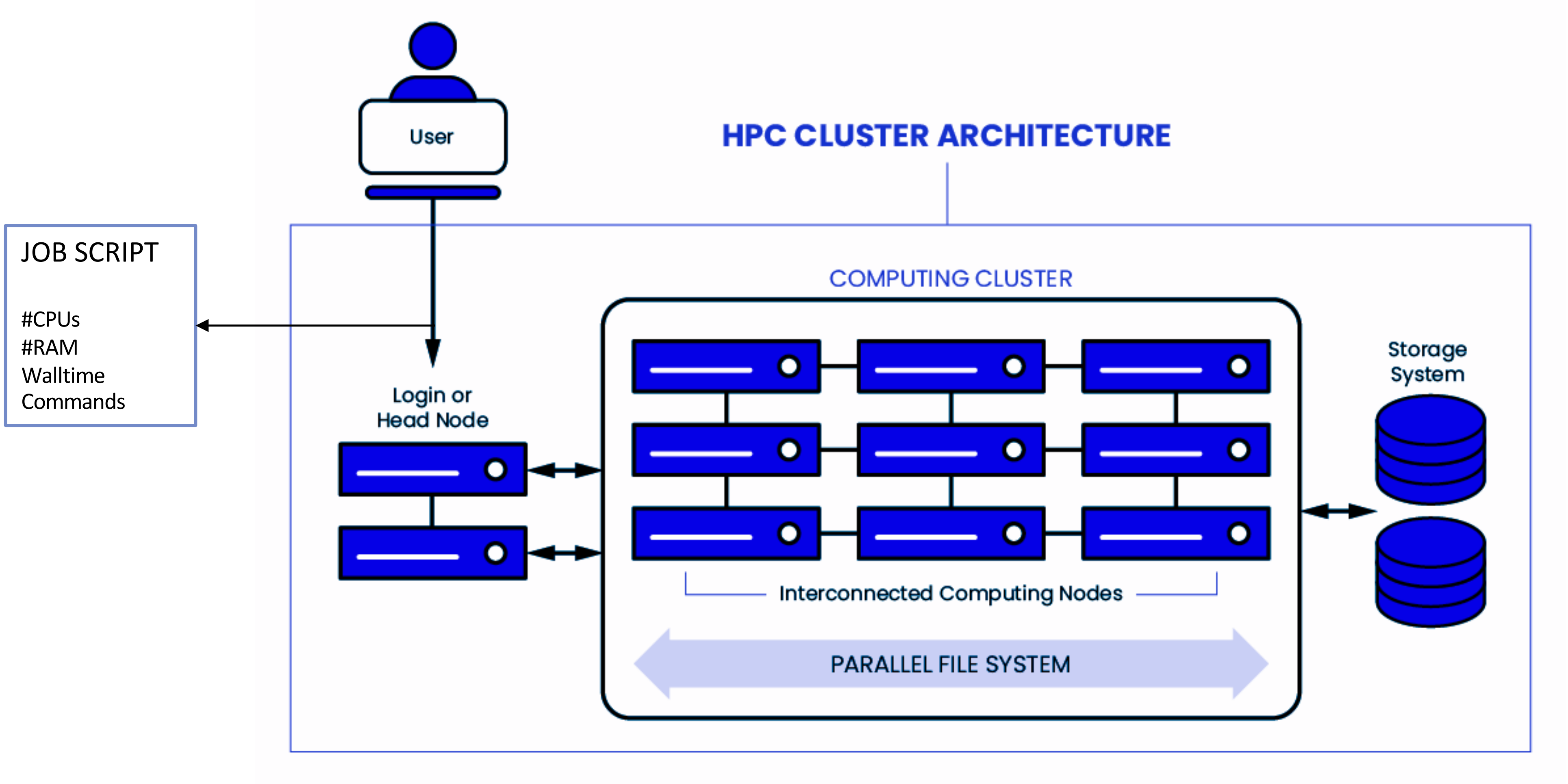
What is High Performance Computing?

Are you using HPC systems?



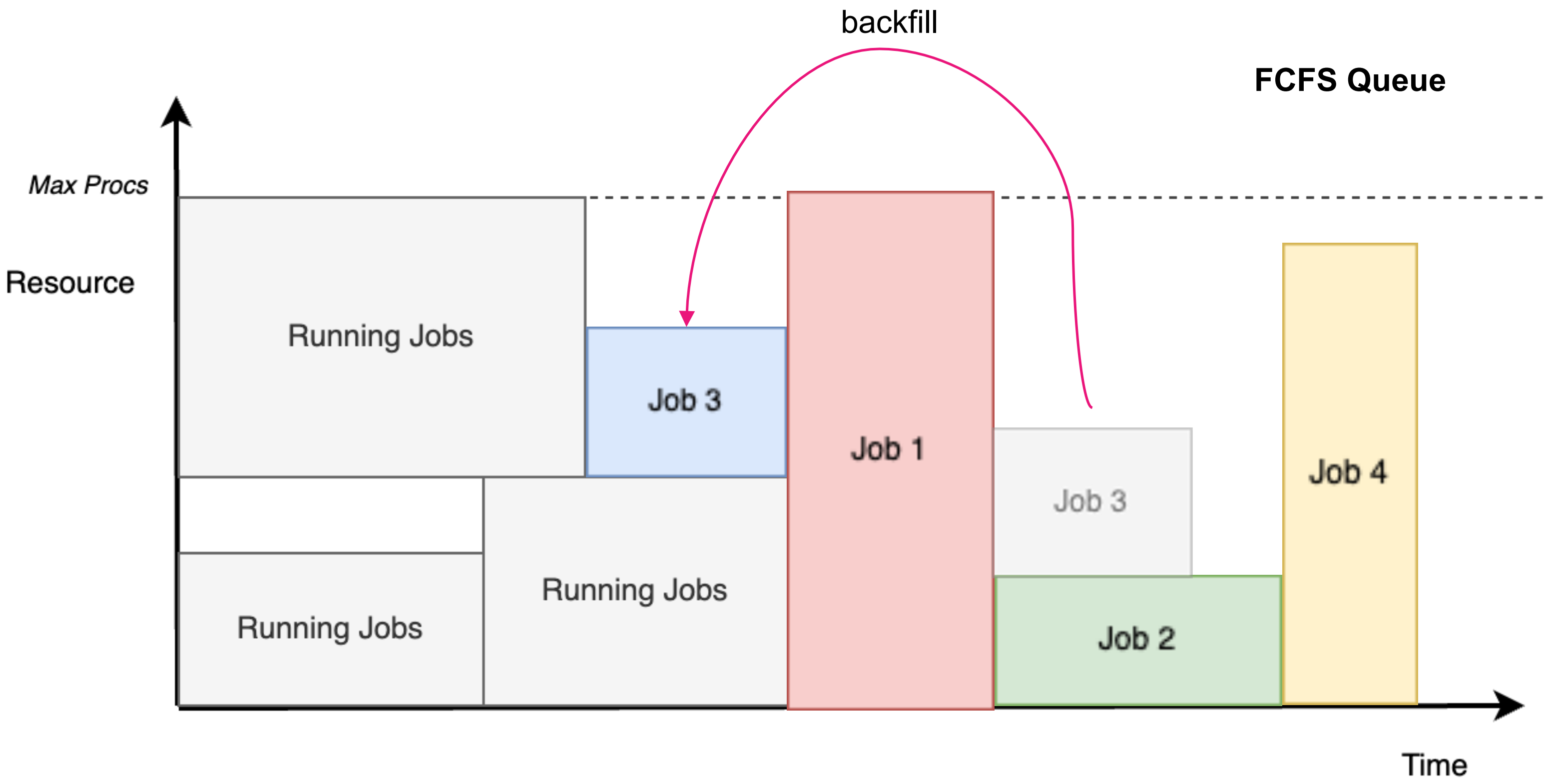
Why Job Runtime Prediction Matters in HPC?

A general workflow on HPC systems



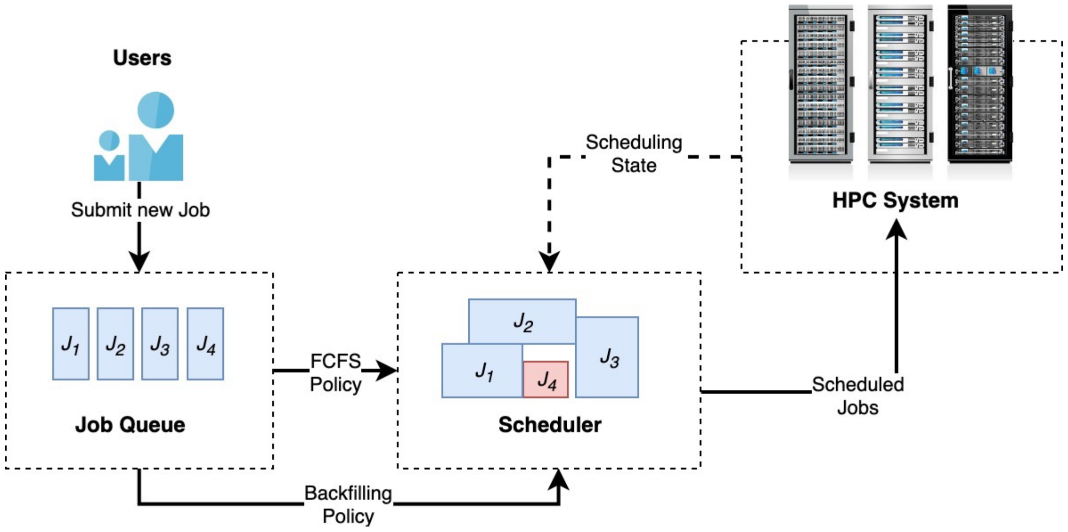
Why Job Runtime Prediction Matters in HPC?

Common scheduling policies such as EASY Backfilling need runtime estimates to perform “backfilling” queued jobs

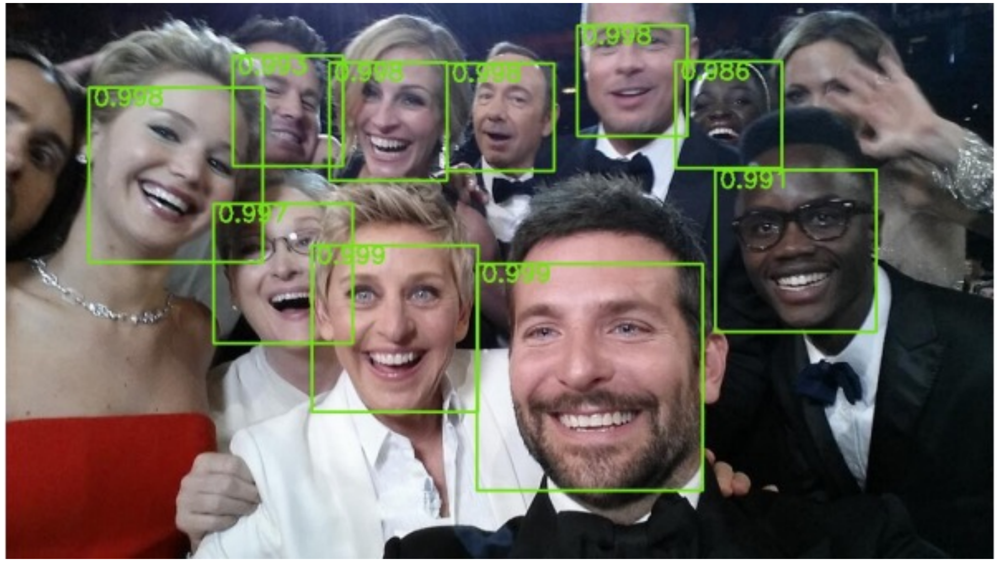


Challenges in Job Runtime Prediction

Is Job Runtime Prediction similar to other common ML problems?



?



Challenges in Job Runtime Prediction

Runtime Prediction is not like other common ML problems

- **Dynamic and Heterogeneous Workloads:**
  - HPC jobs span a wide range of applications:
    - Scientific simulations (e.g., weather forecasting, genomics).
    - AI/ML training workloads.
    - Emerging quantum computing tasks.
  - Workloads vary in computational intensity, memory usage, and I/O demands.

Challenges in Job Runtime Prediction

Runtime Prediction is not like other common ML problems

- **Evolving Hardware Ecosystem**:
  - The transition from **CPU-based** to **GPU-dominated** systems.
  - Increasing adoption of **heterogeneous architectures** (CPU-GPU-QPU hybrids).
  - Hardware upgrades and configurations introduce unpredictability.
- **User-Provided Estimates are Often Inaccurate**:
  - Users tend to **overestimate runtimes** to avoid premature termination.
  - **Underestimations** lead to job failures and resubmissions.

Challenges in Job Runtime Prediction

Runtime Prediction is not like other common ML problems

- **Complex Interdependencies**:
  - Job runtimes depend on multiple factors:
    - System state at submission.
    - Queueing delays.
- **Need for Real-Time Adaptation**:
  - Runtime predictions must adapt to:
    - Changes in system workload patterns.
    - Shifts in user behavior.
    - Variations in job parameters.

Why Single Models Are Not Enough?

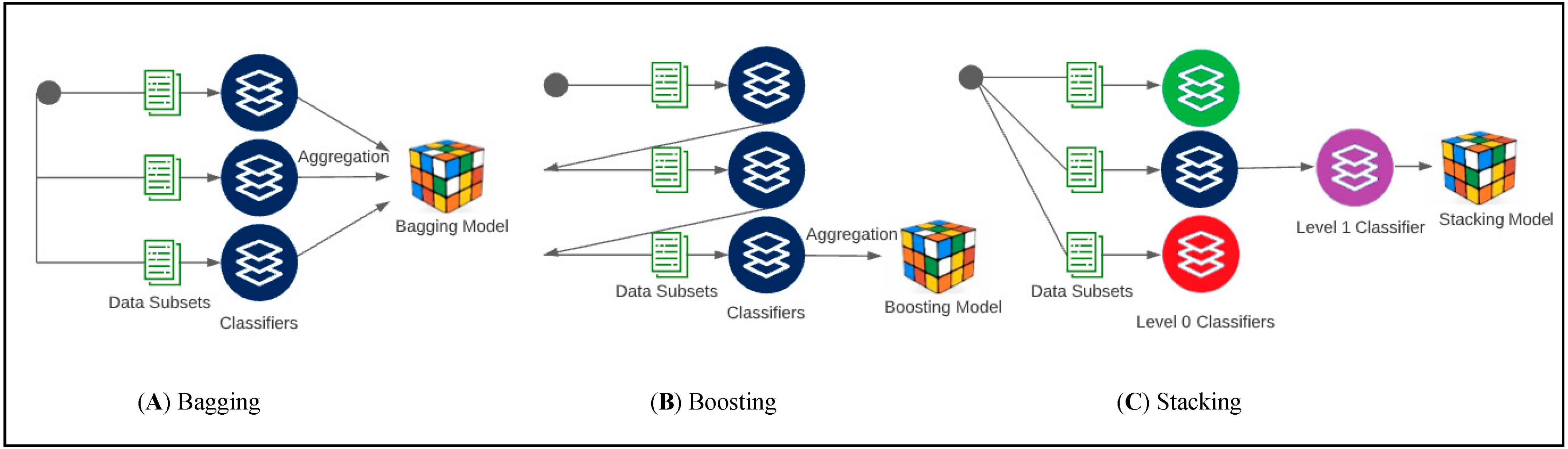
Using a single ML model is a straightforward implementation, but:

- A single model must expand and adapt to capture diverse job characteristics, making it increasingly complex and large.
- Large models require significant computational resources for training and runtime predictions.
- A single model tuned for specific workloads may struggle with unseen patterns or outliers.

Why don’t we combine smaller ML models?

Ensemble Learning

The idea behind ensemble learning is that by combining the strengths of several models, the ensemble can potentially produce better results than any single model.



Ismail, S., El Mrabet, Z., & Reza, H. (2023). An Ensemble-Based Machine Learning Approach for Cyber-Attacks Detection in Wireless Sensor Networks. Applied Sciences, 13(1), 30. https://doi.org/10.3390/app13010030

Research Goals and Objectives

Goal: A robust job runtime prediction model to optimize the resource allocation process in dynamic HPC environments

Objectives:

- Develop an ensemble model that combines multiple predictors to estimate job runtimes accurately.
- Utilize precise runtime predictions to optimize job scheduling.
- Design the predictor to handle a variety of job types and system configurations.
- Allow the model to adapt to changing workload patterns over time.

2. Related Works

History of Job Runtime Prediction in HPC Systems

- **1990s–2000s:** Early Methods with Heuristics and User Estimates.
  - **Mu’alem and Feitelson (2001)** explored the utilization of user runtime estimates in backfilling, showing that their inaccuracies significantly affected scheduling efficiency.
- **2000s–2010s:** Shifted to data-driven methods like regression models for predicting runtime.
  - **Tsafrir et al. (2007)** introduced system-generated runtime predictions, replacing user estimates to improve backfilling performance
- **2010s–Present**: Adopting ML techniques for capturing complex patterns.
  - **Gaussier et al. (2015)** proposed using machine learning for runtime predictions to improve backfilling efficiency, demonstrating significant improvements in scheduling performance.
  - **Fengxian Chen (2023)** demonstrated the ability of Transformer Networks to improve the prediction performance and quality of job scheduling in HPC clusters.

- *Mu’alem, A.W., Feitelson, D.G. (2001): Utilization, predictability, workloads, and user runtime estimates in scheduling the IBM SP2 with backfilling. IEEE Transactions on Parallel and Distributed Systems, 12(6), 529–543.*
- *Tsafrir, D., Etsion, Y., Feitelson, D.G. (2007): Backfilling using system-generated predictions rather than user runtime estimates. IEEE Transactions on Parallel and Distributed Systems, 18(6), 789–803.*
- *Gaussier, E., Glesser, D., Reis, V., Trystram, D. (2015): Improving backfilling by using machine learning to predict running times. In SC '15: International Conference for High-Performance Computing, Networking, Storage and Analysis (pp. 1–10).*
- *Chen, F. Job runtime prediction of HPC cluster based on PC-Transformer. J Supercomput 79, 20208–20234 (2023). https://doi.org/10.1007/s11227-023-05470-2*

Ensemble Learning for Job Runtime Prediction

Many works have utilized ensemble learning for job runtime prediction:

- **Tanash et al. (2021)** combined multiple ML models to predict resource demands in SLURM-based HPC systems.
- **Bai et al. (2022)** proposed a Multivariate Resource Demand Prediction (MRDP) model using ensemble learning with a random forest meta-learner.
- **Ramachandran et al. (2024)** integrated ML models and genetic algorithms for runtime prediction.

However:

- Existing ensemble models lack real-time feedback mechanisms to adapt to workload changes and system upgrades.
- Limited studies evaluate the direct impact of runtime prediction accuracy on scheduling efficiency.

- *Tanash, M., Yang, H., Andresen, D., Hsu, W. (2021): Ensemble prediction of job resources to improve system performance for SLURM-based HPC systems. In PEARC '21.*
- *Bai, Y., Guo, Y., Zhang, H., Wang, J., Chen, J. (2022): An ensemble learning-based HPC multi-resource demand prediction model for hybrid clusters. In ICCSMT 2022.*
- *Ramachandran, S., Jayalal, M., Vasudevan, M., Das, S., Jehadeesan, R. (2024): Combining machine learning techniques and genetic algorithms for predicting runtimes of HPC jobs. Applied Soft Computing, 165, 112053.*

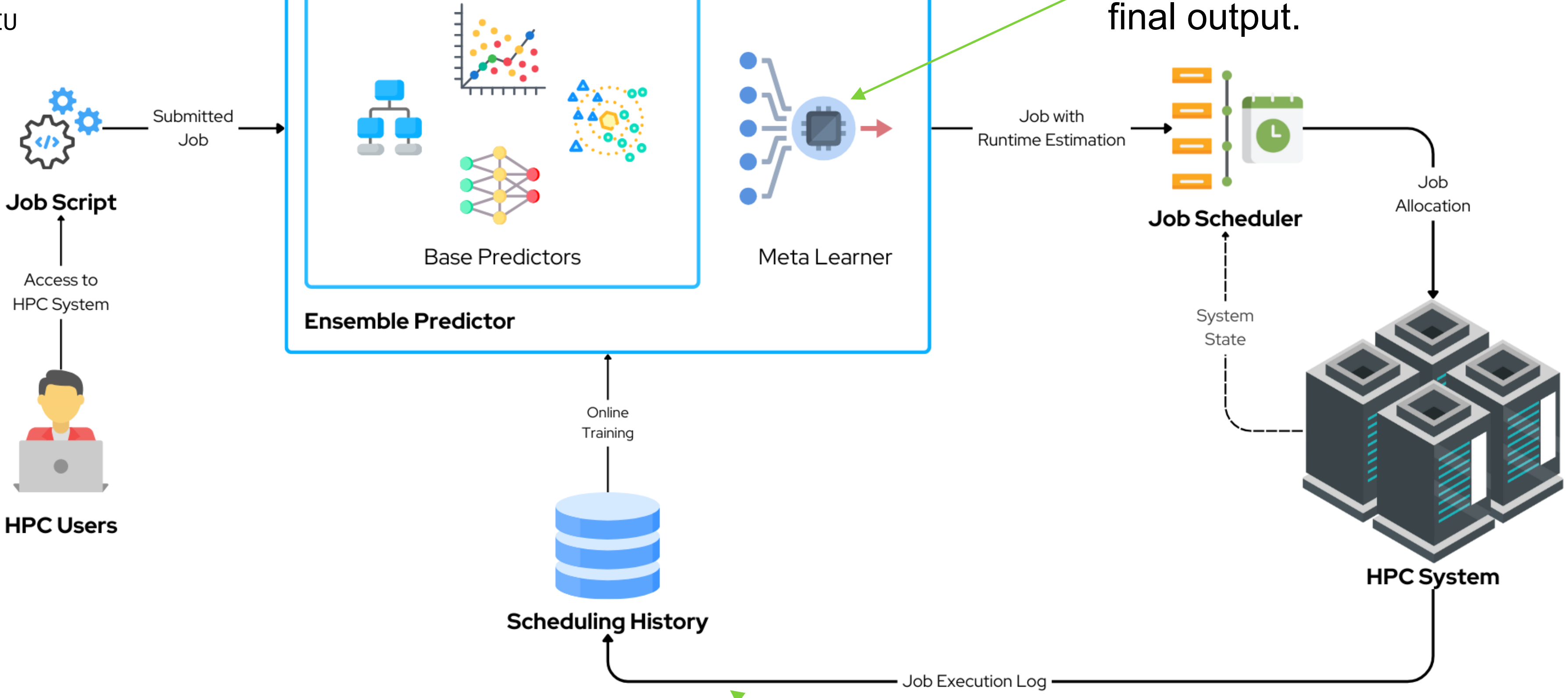
3. Proposed Method

JREP Overview

Base models: kNN, RF, SVM, DNN – each captures different data dimensions.

Job Info:

- Requested Resources
- Execution IU
- User ID
- Group ID
- Queue



Meta-learner: Combines base predictions for final output.

Feedback loop: Continuous learning ensures adaptability to dynamic workloads.

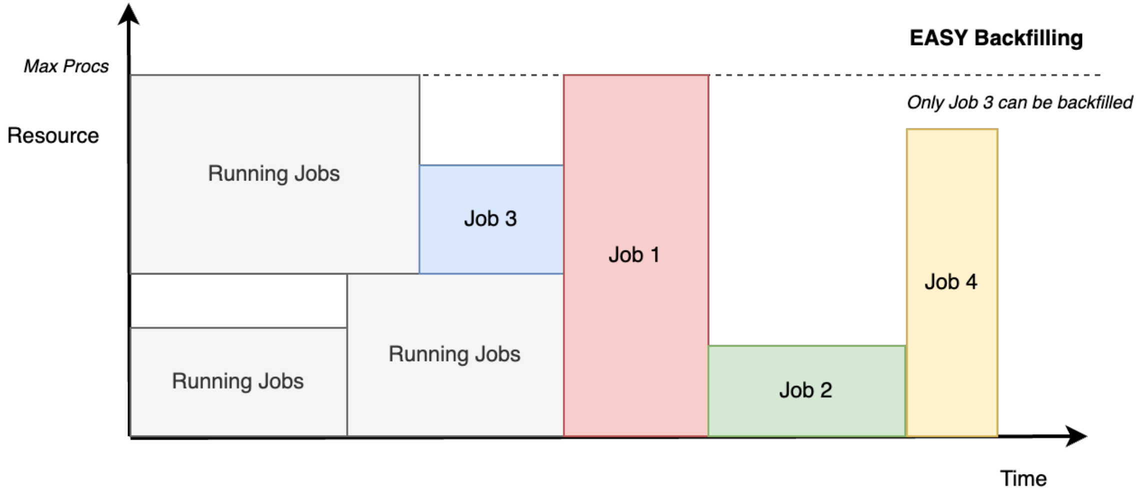
Job **R**untime **E**nsemble **P**redictor

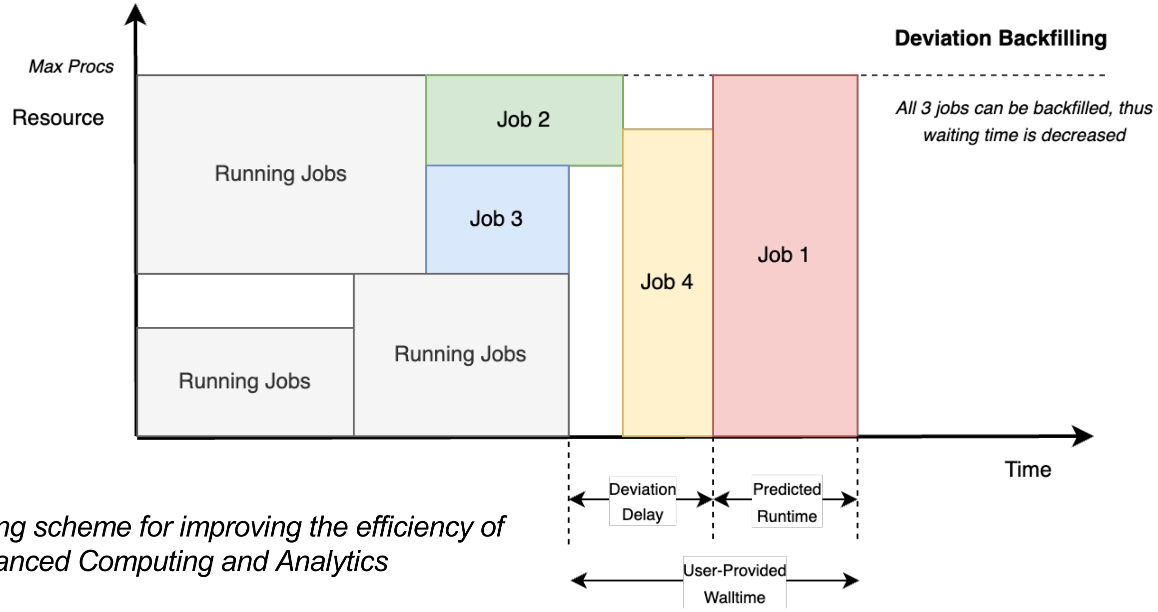
4. Experimental Results

Evaluating JREP with Deviation Backfilling

Deviation Backfilling extends traditional backfilling by accounting for **runtime prediction deviations** to improve scheduling decisions.

Better Prediction = More Space to Backfill Queued Jobs

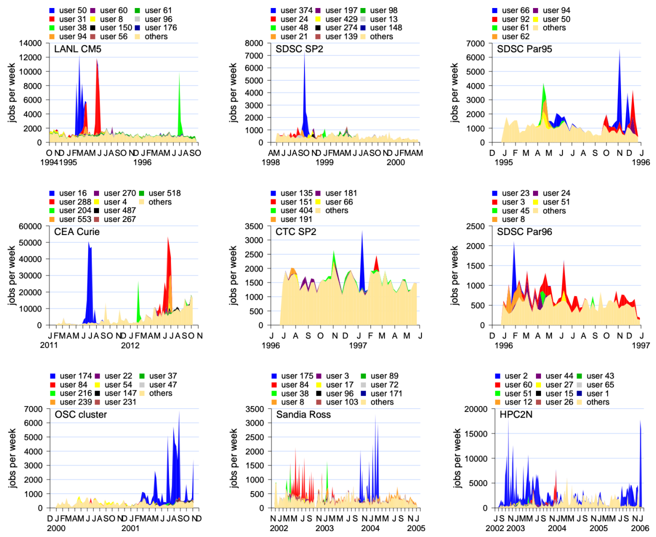




Le Hai, T.H., Duy, K.N., Manh, T.N., Hoang, D.M., Thoai, N.: Deviation backfilling: a robust backfilling scheme for improving the efficiency of job scheduling on high performance computing systems. In: 2023 International Conference on Advanced Computing and Analytics (ACOMPA), pp. 32–37 (2023)

Real Workload Datasets

Scheduling Datasets: **Parallel Workload Archive (PWA)** + SuperNode-XP



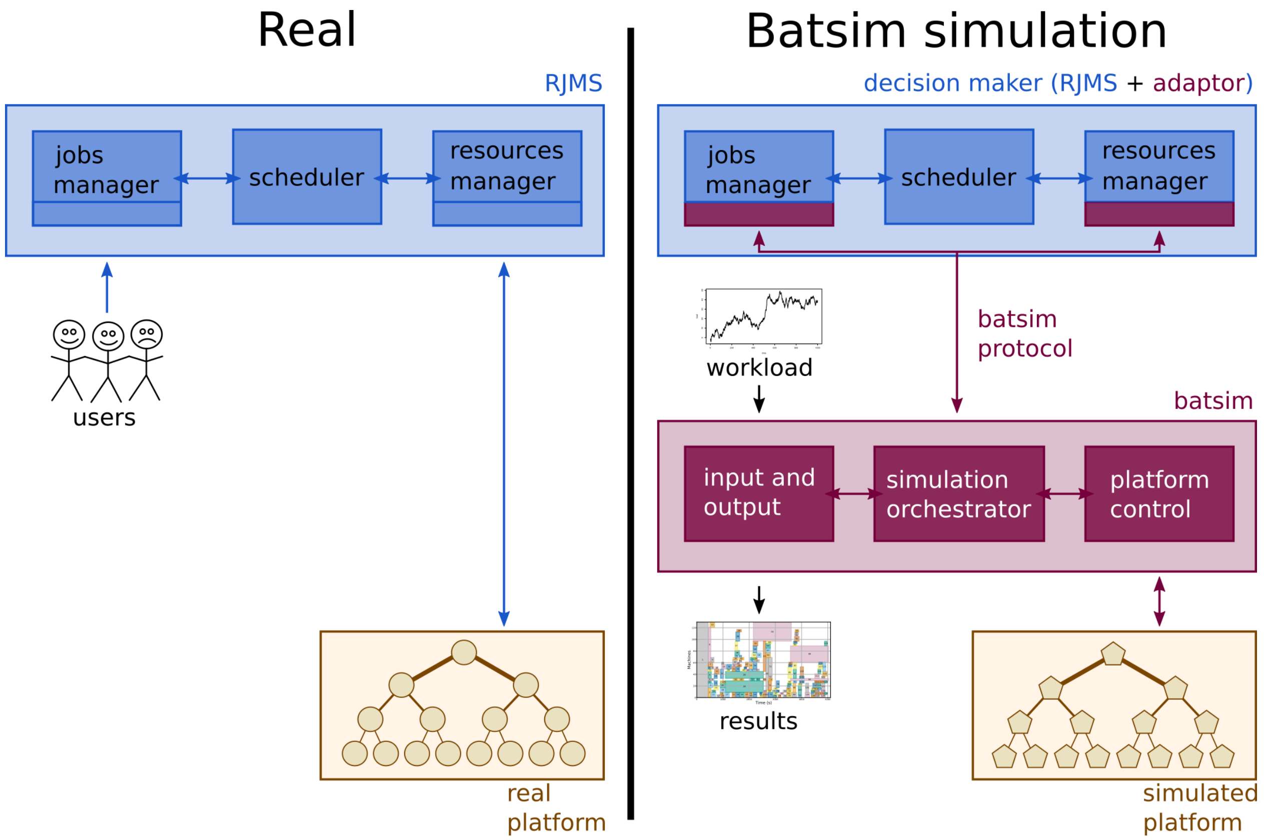
Workload Log	Time	#Jobs	#Nodes
SDSC-DS-2004	Mar 2004 - April 2005	96,069	171
ANL-Intrepid-2009	Sep 2009 - May 2010	68,936	640
SuperNode-XP-2017	May 2017 - March 2021	19,124	24

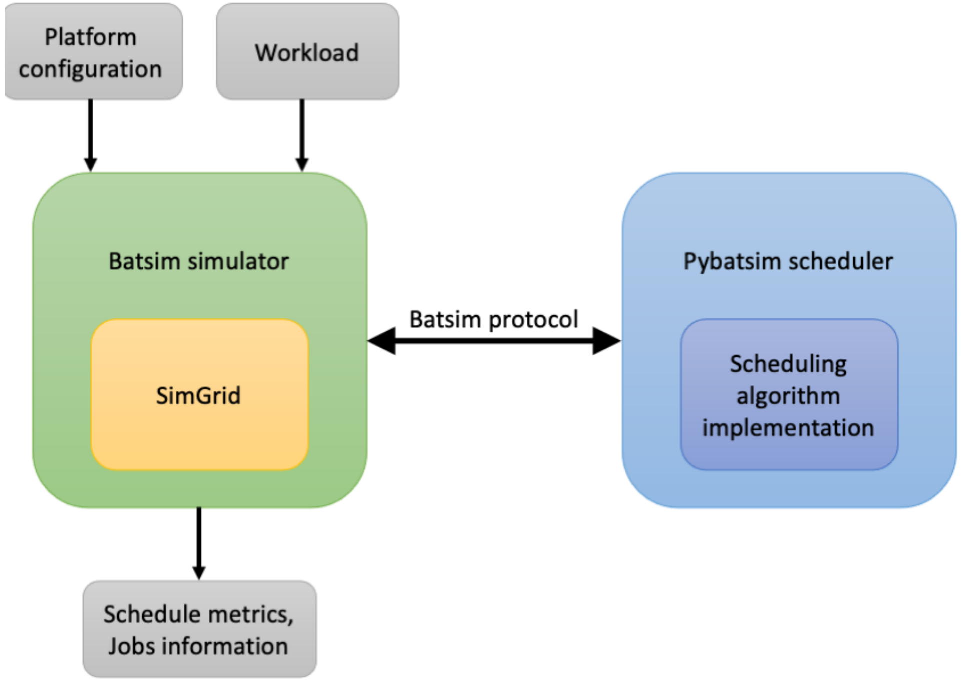
Dror G. Feitelson, Dan Tsafrir, David Krakov, Experience with using the Parallel Workloads Archive, Journal of Parallel and Distributed Computing, Volume 74, Issue 10, 2014

https://www.cs.huji.ac.il/labs/parallel/workload

Scheduling Simulation

Simulation tool: **Batsim with PyBatsim**



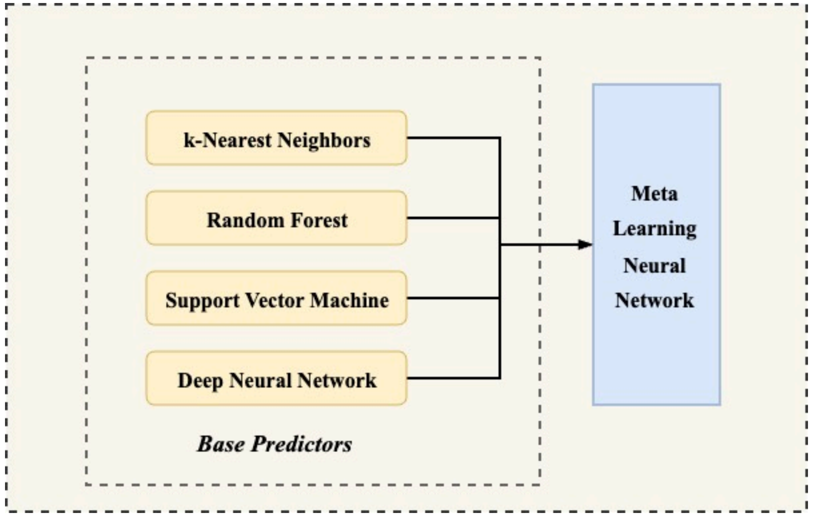


https://batsim.readthedocs.io/en/latest/

Base Predictors Setup

Implemented by scikit-learn in Python

Method	Parameters
RF	n_estimators = 500, max_depth = None, min_samples_split = 2, max_features = 'auto', bootstrap = True
KNN	n_neighbors = 5, weights = 'distance', algorithm = 'auto', leaf_size = 30
SVM	kernel = 'rbf', C = 2.0, epsilon = 0.1, gamma = 'scale'
DNN	hidden_layer_sizes = (64, 128, 64), activation = 'relu', solver = 'adam', max_iter = 1000, learning_rate_init = 0.001
Meta Learner (DNN)	hidden_layer_sizes = (64, 128, 32), activation = 'relu', solver = 'adam', max_iter = 1000, learning_rate_init = 0.001



https://batsim.readthedocs.io/en/latest/

Scheduling Performance Result

Scheduling Policy	SDSC-DS-2004	ANL-Intrepid-2009	SuperNode-XP-2017
EASY Backfilling	1098.2s	4982.7s	160158.1s
Deviation Backfilling (w-kNN)	**967.5s**	4149.8s	140861.1s
Deviation Backfilling (JSEP)	979.2s	**3875.6s**	**132803.9s**

Average Job Wait Time *(Lower is better)*

Scheduling Policy	SDSC-DS-2004	ANL-Intrepid-2009	SuperNode-XP-2017
EASY Backfilling	6.22	10074.33	7.55
Deviation Backfilling (w-kNN)	**5.69**	7343.79	6.19
Deviation Backfilling (JSEP)	5.81	**5759.91**	**5.97**

Average Job Slowdown *(Lower is better)*

Ensemble Prediction Performance

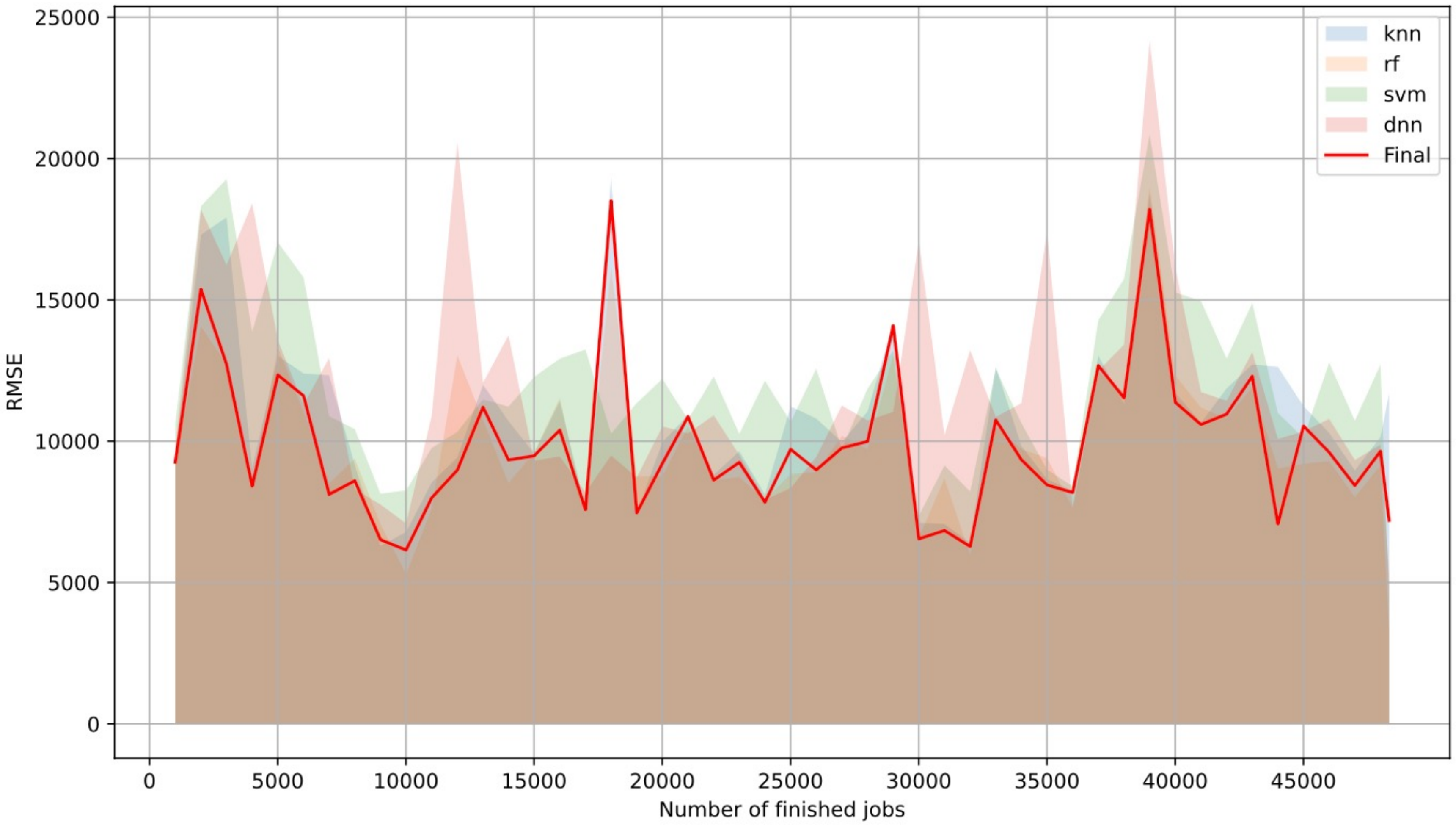
Configuration	ANL-Intrepid-2009	SDSC-DS-2004	SuperNode-XP-2017
JREP	**7902.91**	10280.30	**152990.48**
JREP w/o DNN	8144.56	**10202.49**	158158.78
JREP w/o RF	8282.08	11003.38	166371.42
JREP w/o kNN	8724.68	10439.66	176607.71
JREP w/o SVM	8251.55	10388.23	153210.82

RSME Comparison Across Datasets *(Lower is better)*

Configuration	ANL-Intrepid-2009	SDSC-DS-2004	SuperNode-XP-2017
JREP	**0.2359**	0.2095	**-0.1852**
JREP w/o DNN	0.1885	**0.2214**	-0.2666
JREP w/o RF	0.1608	0.0944	-0.4015
JREP w/o kNN	0.0687	0.1848	-0.5793
JREP w/o SVM	0.1670	0.1928	-0.1886

R^2 Comparison Across Datasets *(Higher is better)*

Ensemble Prediction Performance



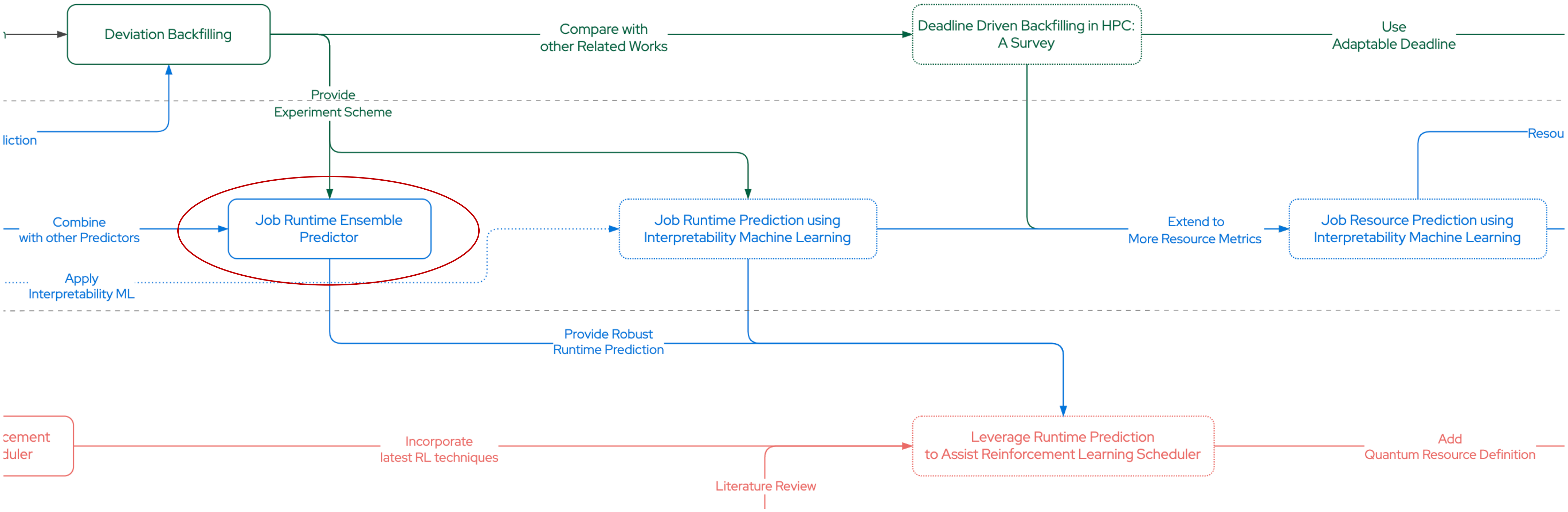
RMSE for Every finished 1000 Jobs for Each Predictor (SDSC-DS-2004)

5. Conclusions

Remarks

- Accurate runtime prediction is essential for optimizing resource allocation and scheduling in HPC systems.
- We introduced a novel approach called JREP to provide a scalable, adaptable, and accurate prediction framework that can be directly integrated into scheduling mechanisms.
- We hope that our contribution can be a foundation for further exploration of ensemble learning in dynamic, heterogeneous HPC environments.

What's next

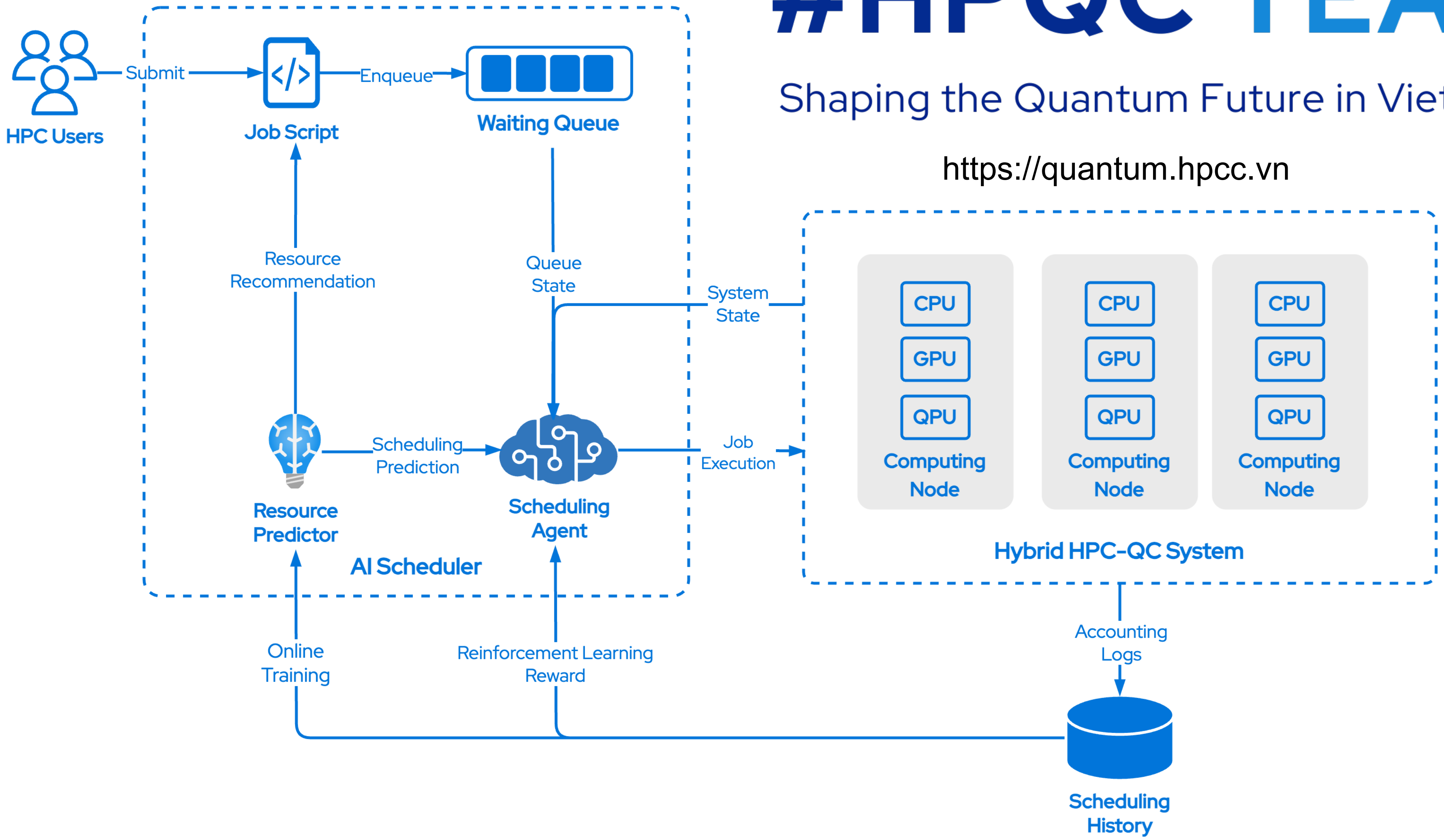


Deviation Backfilling
Compare with
other Related Works
Deadline Driven Backfilling in HPC:
A Survey
Use
Adaptable Deadline
Provide
Experiment Scheme
liction
Resou
Combine
with other Predictors
Job Runtime Ensemble
Predictor
Job Runtime Prediction using
Interpretability Machine Learning
Extend to
More Resource Metrics
Job Resource Prediction using
Interpretability Machine Learning
Apply
Interpretability ML
Provide Robust
Runtime Prediction
cement
duler
Incorporate
latest RL techniques
Leverage Runtime Prediction
to Assist Reinforcement Learning Scheduler
Add
Quantum Resource Definition
Literature Review

#HPQC TEAM

Shaping the Quantum Future in Viet Nam

https://quantum.hpcc.vn



The 11th International Conference on Future Data and Security Engineering

FDSE 2024

VIETNAMESE-GERMAN UNIVERSITY, BEN CAT, BINH DUONG, VIETNAM

November 27-29, 2024

Springer CCIS VGU Vietnamese-German University LIT Secure and Correct Systems Lab FAW SPRINGER NATURE

Thanh Hoang Le Hai
thanhhoang@hcmut.edu.vn

High Performance Computing Lab
Faculty of Computer Science & Engineering, HCMUT, VNU-HCM

Thank you

www.hpcc.vn hpcc@hcmut.edu.vn