

# Kubernetes Resource Scaling via Batch Node Conversion on the Anvil Supercomputer

Erik Gough  
Purdue University  
West Lafayette, USA  
goughes@purdue.edu

LJ Lumas  
Purdue University  
West Lafayette, USA  
dlumas@purdue.edu

## Abstract

The flexibility and scalability of the Kubernetes container orchestration system have made it a staple in enterprise cloud computing environments. Recently, research institutions have increasingly adopted Kubernetes for container-based scientific computing, often deploying it alongside HPC resources to support applications such as science gateways and JupyterHub instances. However, this Kubernetes infrastructure frequently has fewer resources compared to dedicated HPC systems. When demand for Kubernetes resources exceeds their capacity, such as during large workshops or training sessions, it can become challenging to accommodate all participants.

This lightning talk will explore our approach to reallocating compute nodes in an HPC system managed by Slurm, converting them from "batch" nodes to "cloud" nodes within a Kubernetes resource. We developed and implemented these methods on the Anvil ACCESS resource to support large-scale educational and training workshops. By temporarily shifting batch nodes to cloud nodes, we overcame capacity limitations on Anvil's Kubernetes infrastructure, enabling one workshop to scale up to 75 computing sessions, a 3.5x increase over what was possible with their allocation. We will give an overview of Anvil's xCAT + masterless puppet configuration management stack and introduce a script that facilitates the conversion of HPC batch nodes to Anvil Kubernetes nodes and back, providing flexible resource management through command-line options. While each institution's software stack is unique, our experience and guidelines offer a foundation that others can adapt to achieve similar success in their own environments.

## CCS Concepts

• Information systems → Computing platforms; • Computer systems organization → Cloud computing.

## Keywords

Cloud computing, High performance computing, Kubernetes

## ACM Reference Format:

Erik Gough and LJ Lumas. 2024. Kubernetes Resource Scaling via Batch Node Conversion on the Anvil Supercomputer. In *Proceedings of ?? (HPC-SYSPROS 24)*. ACM, New York, NY, USA, 1 page. <https://doi.org/XXXXXXX.XXXXXXX>

## 1 Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 2005632. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

---

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

HPCSYSPROS 24, November 22, 2024, Atlanta, GA

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.  
ACM ISBN 978-x-xxxx-xxxx-x/YY/MM  
<https://doi.org/XXXXXXX.XXXXXXX>