Lube: Mitigating Bottlenecks in Wide Area Data Analytics

UNIVERSITY OF TORONTO

Hao Wang and Baochun Li, Department of Electrical and Computer Engineering, University of Toronto

Introduction

Flink Spark hadoop TACHYON

DC
- Master
- Namenode
- Workers
- Datanodes

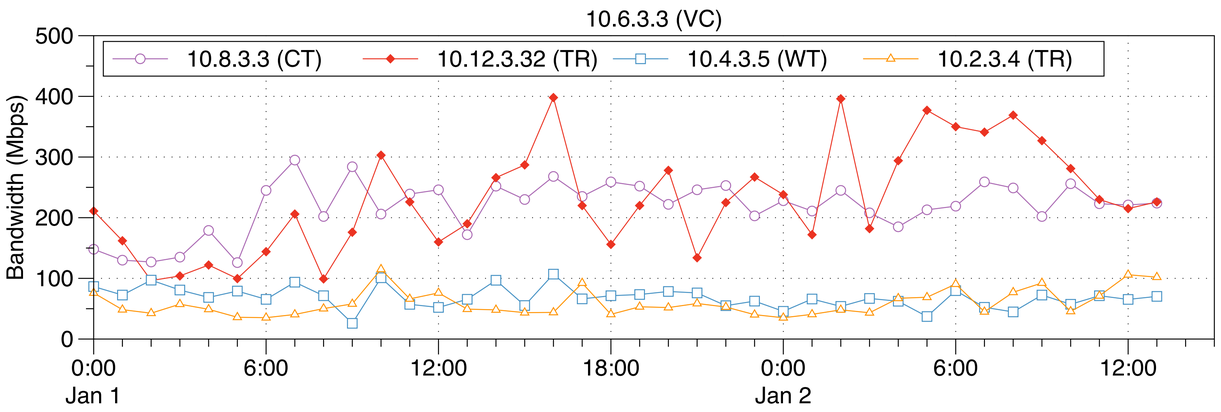
DC#1 (Master, Namenode) | DC#2 ... DC#n (Workers ... Workers, Datanodes ... Datanodes)

From single datacenter to geo-distributed datacenters

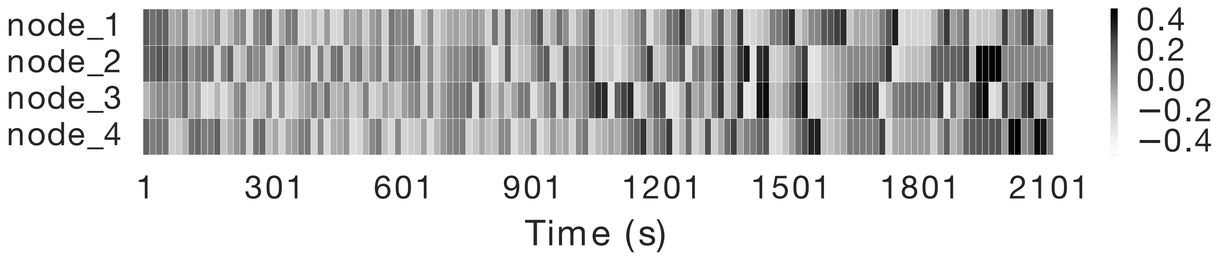
- Data Volume
- User Distribution
- Regulation Policy

→ Widely Shared Resources → Fluctuating Available Provision

→ Distributed Runtime Environment → Heterogenous Utilizations



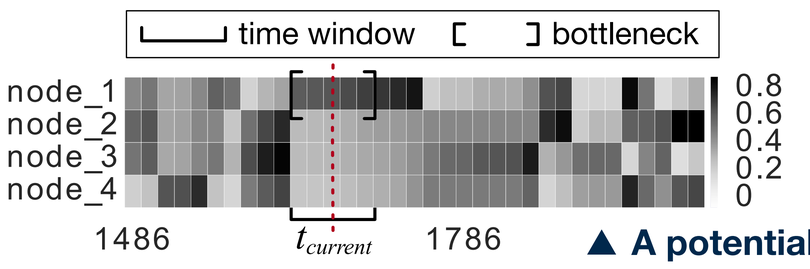
Fluctuating WAN bandwidths ▲



▼ Heterogenous memory utils

Bottlenecks

Resource Heterogeneity + Performance Fluctuation → **Bottlenecks**: Any time; Any nodes; Any resources



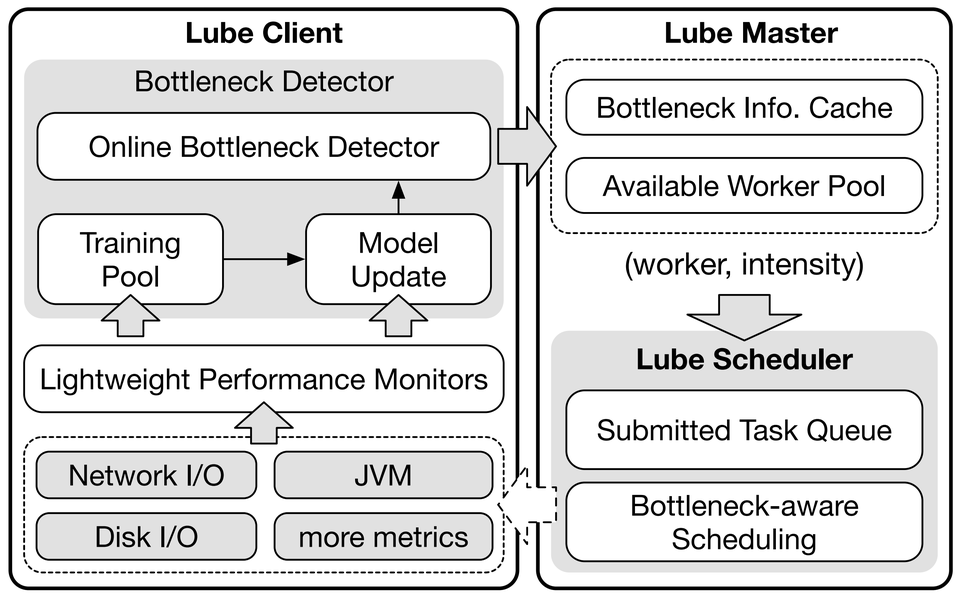
▲ A potential bottleneck in memory

Mitigating Bottlenecks at Runtime

- ✓ How to detect bottlenecks?
- ✓ How to overcome scheduling delay?
- ✓ How to schedule with an awareness of bottlenecks?

Lube Overview

- Performance Monitors
- Bottleneck Detecting Module
- Bottleneck-Aware Scheduler

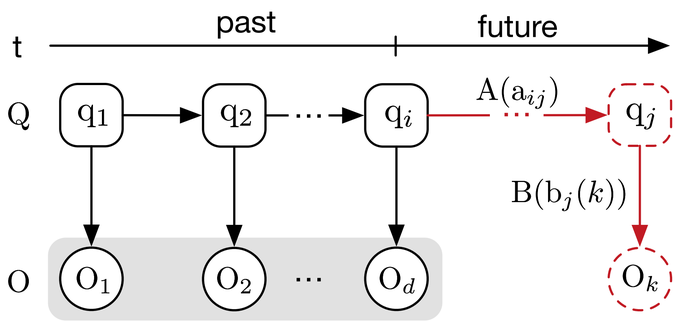


Detecting Bottlenecks

$$y_t = \theta_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

Historical States → input → Autoregressive (AR) + Moving Average(MA) → output → Historical States

▲ ARIMA

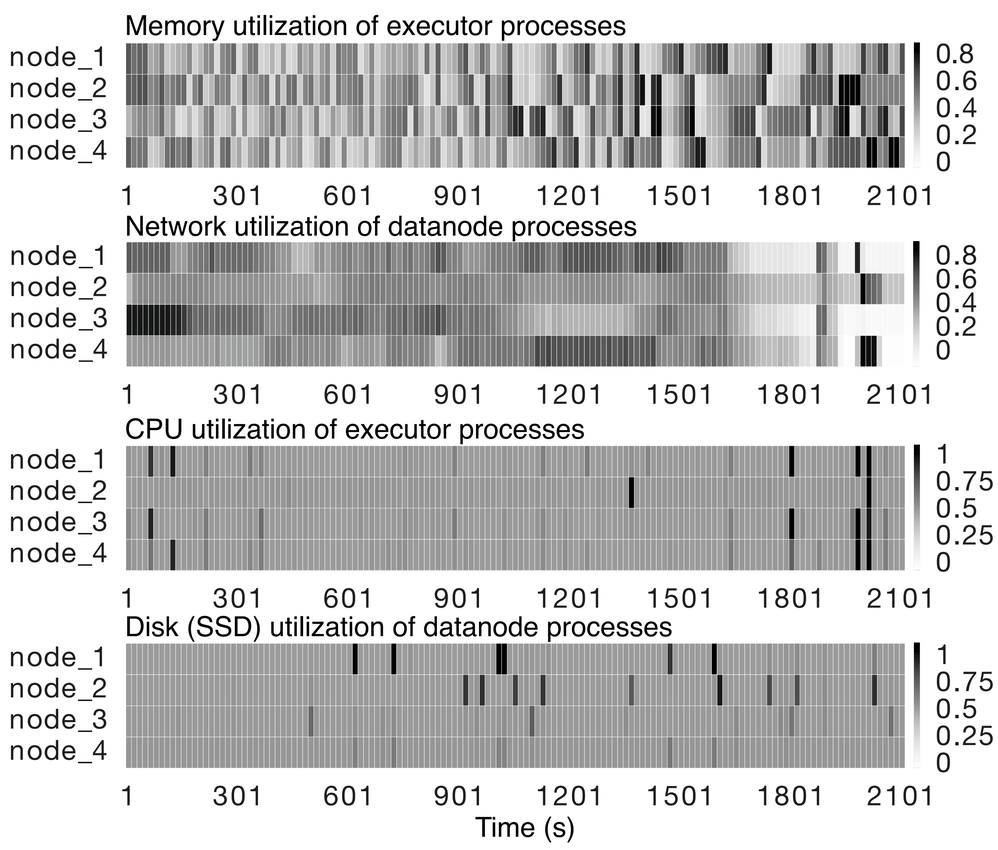


▼ SlidHMM

O: Hidden states
Q: Observation states
A: Emission probability
B: Transition probability

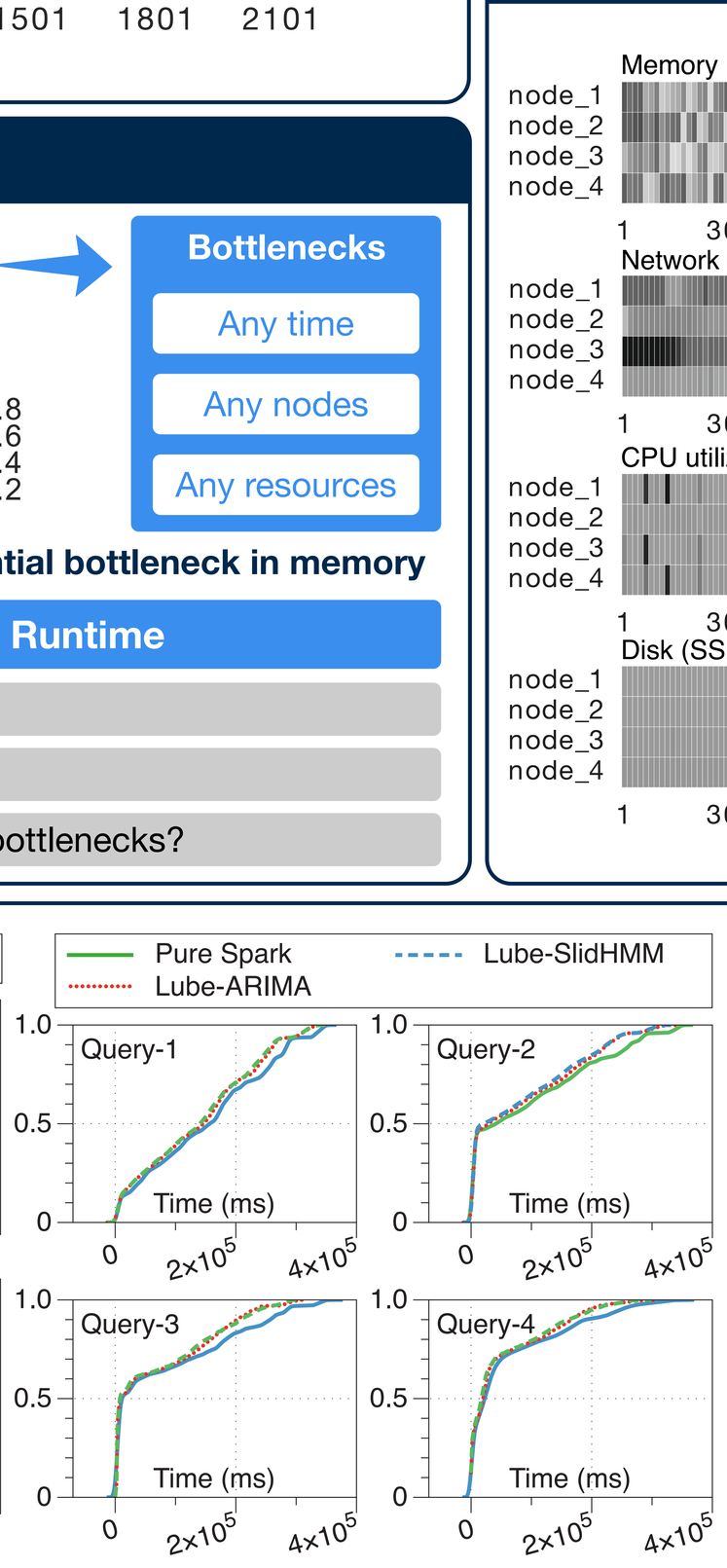
{time_stamp: mem, net, cpu, disk}

Bottleneck-Aware Scheduling



Data-locality → Data-locality + Bottlenecks

Evaluation



Acknowledgement

This work is supported by a research contract with Huawei Technologies Co. Ltd. and an NSERC Collaborative Research and Development (CRD) grant. We would like to thank the HotCloud anonymous reviewers for their valuable comments.