Multi-Node Spot Instances Availability Score Collection System

Sungkyu Cheon, Kyumin Kim, Kyunghwan Kim, Moohyun Song and Kyungyong Lee

Success rate variation when single-node SPS is 3 **1.0** _⊤ Ratio 8.0 quest 0.0 Fulfillment F 0.2 5.0

Number of Requested Nodes

Target Node 1

2012012012012012012022022022023

Released Year

Re

2.50

1.75

1.50

1.25

1.00

SdS 2.00

What Is Spot Dataset

Spot instances are cheap and unstable

- Up to 90% cheaper unused cloud resources
- Could be interrupted or not-fulfilled

Cloud vendors provide their own dataset

- Spot Price Dataset (On-demand, Spot, Savings)
- Spot Availability Dataset (SPS)

Spotlake provides single-node SPS data

- Single-node SPS == 3, High success at 1 node, Drop at 50 nodes
- → Single-node SPS is unreliable for large deployments

201201201201201201202022022023

Released Year

Goal

Analysis Single-node SPS

- Can the single-node data useful for large deployments?

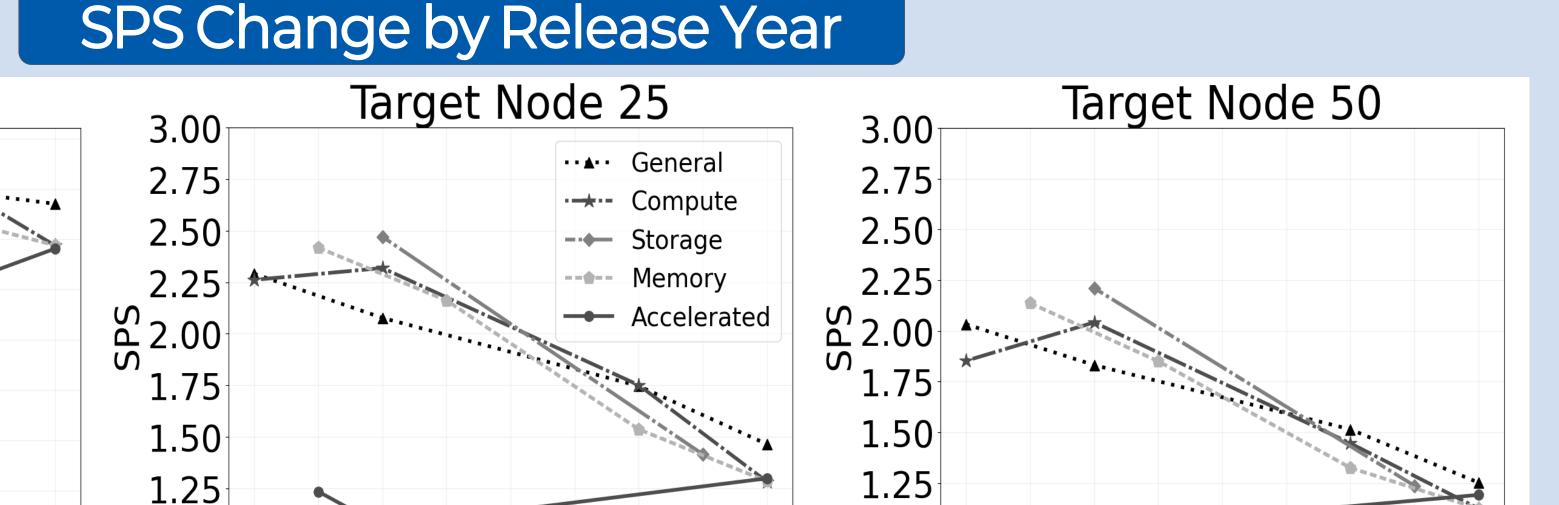
Collect Multi-node SPS

- Can we collect a multi-node availability collection system?
- Can we provide multi-node SPS dataset?

Analysis Multi-node SPS

- How does multi-node SPS differ?

Data Analysis



1.00

Newer Instance Type → Lower SPS

Accelerated Computing → Higher SPS → High cost & Software Complexity

➤ Release year and category influence multi-node availability

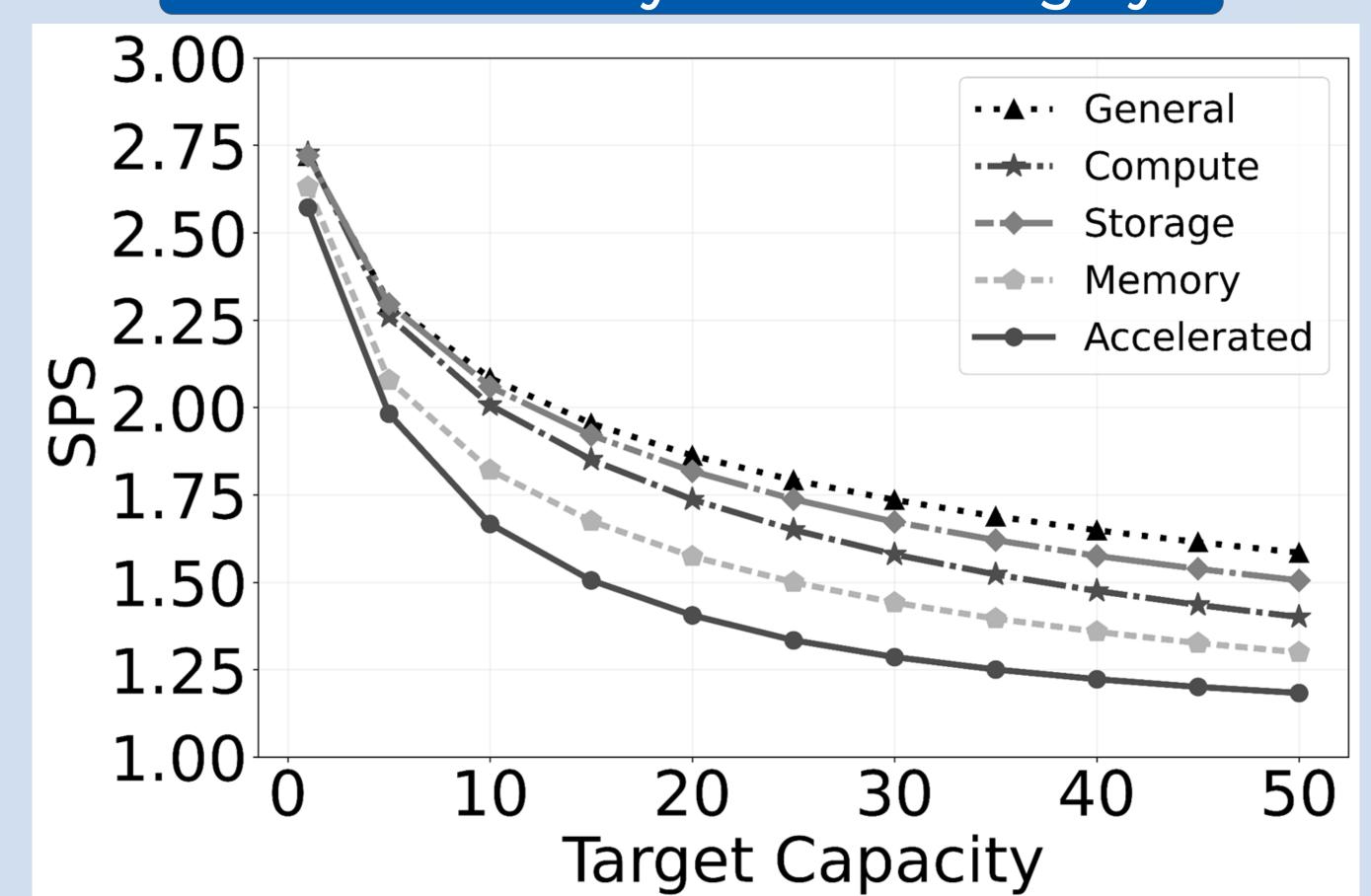
SPS variations by instance category

1.25

1.00

201201201201201201202022022023

Released Year



Node Count A, SPS v

Accelerated Computing shows the sharpest decline (high GPU demand)

> Instance category is critical for multi-node deployments

Spot Instance Dataset **Archive Service Online**

- Periodic SPS queries to AWS (every 10 minutes)
- Collects availability for single- and multi-node
- Enhanced Spotlake with real-time/historical multi-node views

