

# Pricing Games for Hybrid Object Stores in the Cloud: Provider vs. Tenant

Yue Cheng<sup>★</sup>, M. Safdar Iqbal<sup>★</sup>, Aayush Gupta<sup>†</sup>, Ali R. Butt<sup>★</sup>

<sup>★</sup>*Virginia Tech* & <sup>†</sup>*IBM Research – Almaden*



# Cloud-based object stores

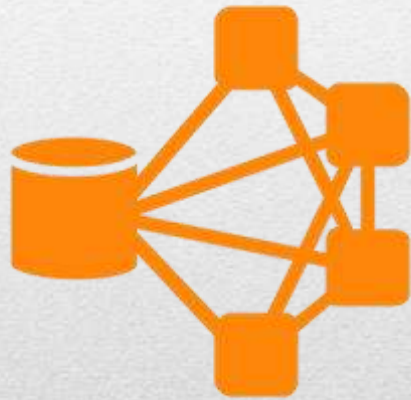


# Data analytics on object stores

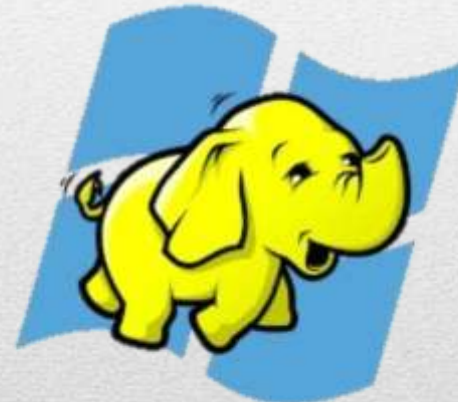




# Data analytics on object stores



Amazon Web Service  
Elastic MapReduce

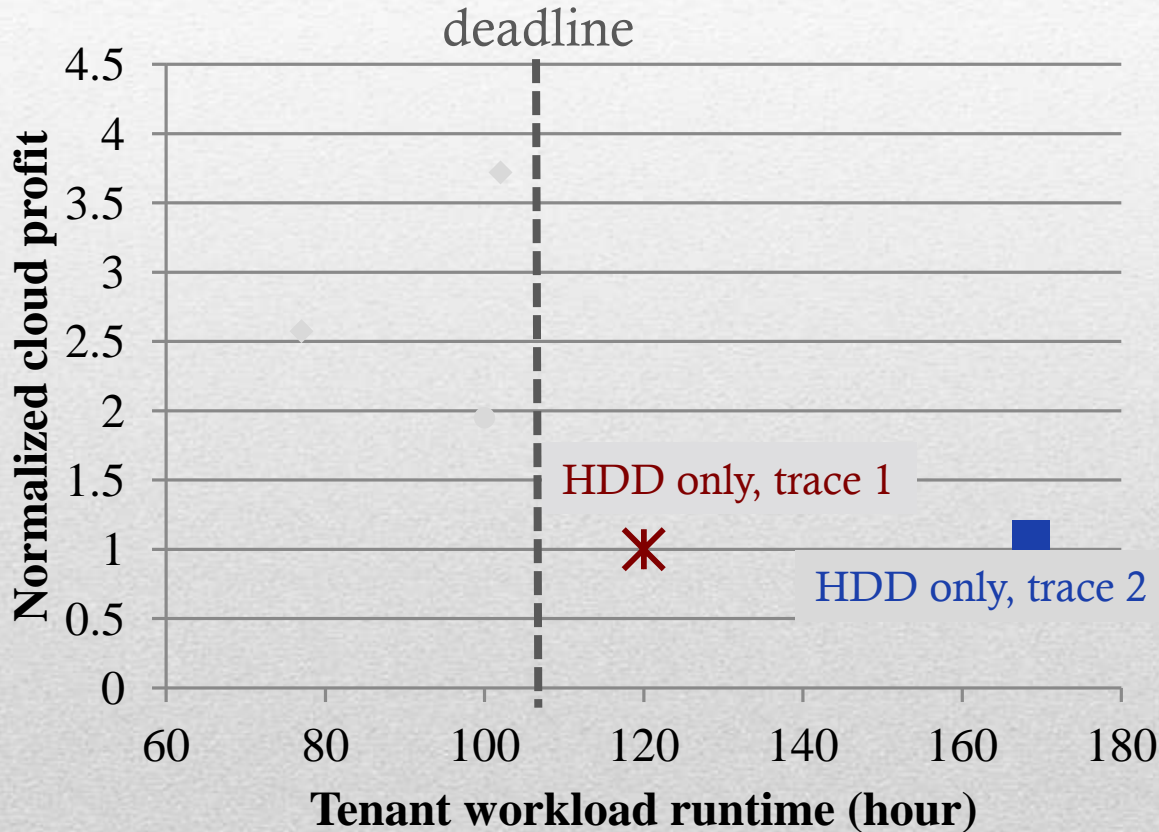


Microsoft Azure  
HDInsight

# Cloud-based object stores



# HDD-based object store?



HDD price: \$0.0011/GB/day

Trace 1: 12TB input 5TB output

Trace 2: 18TB input 8TB output



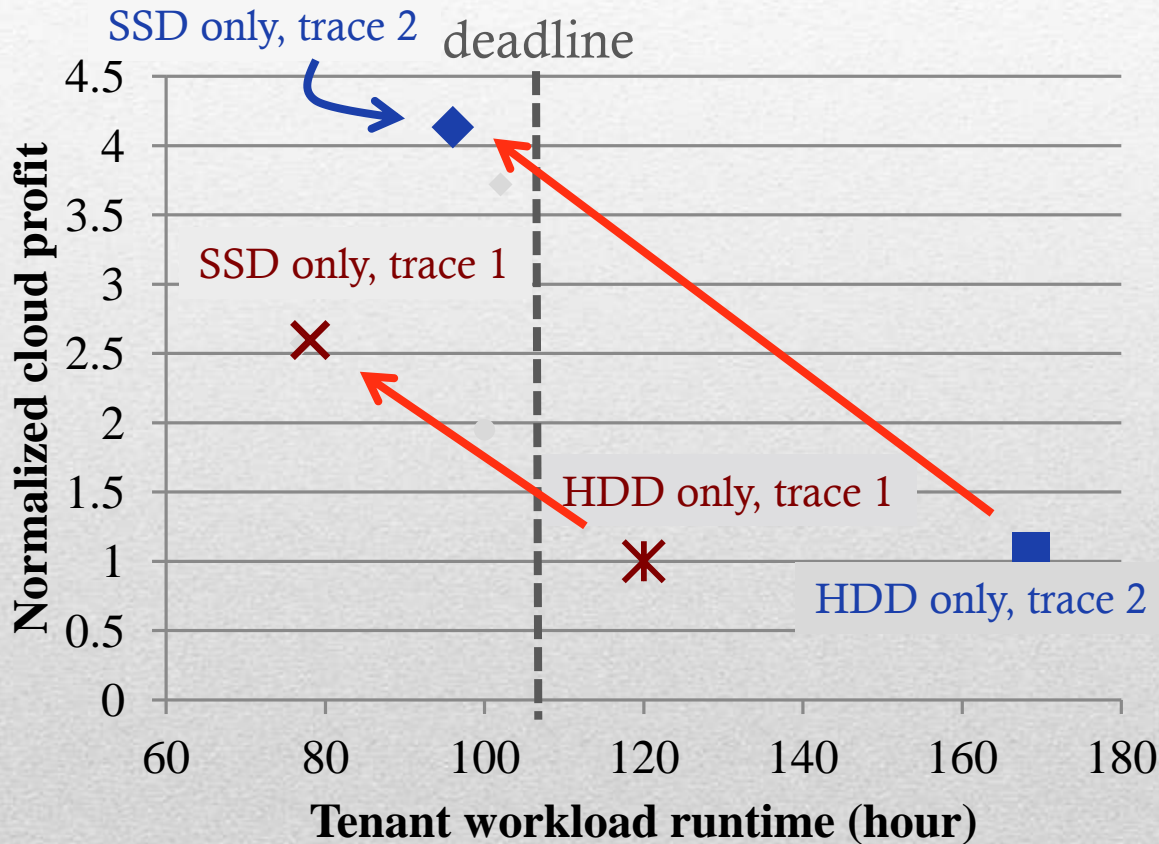
Tenants cannot meet deadline



Provider get low profit



# SSD-based object store?



HDD price: \$0.0011/GB/day  
SSD price: \$0.0044/GB/day

Trace 1: 12TB input 5TB output  
Trace 2: 18TB input 8TB output

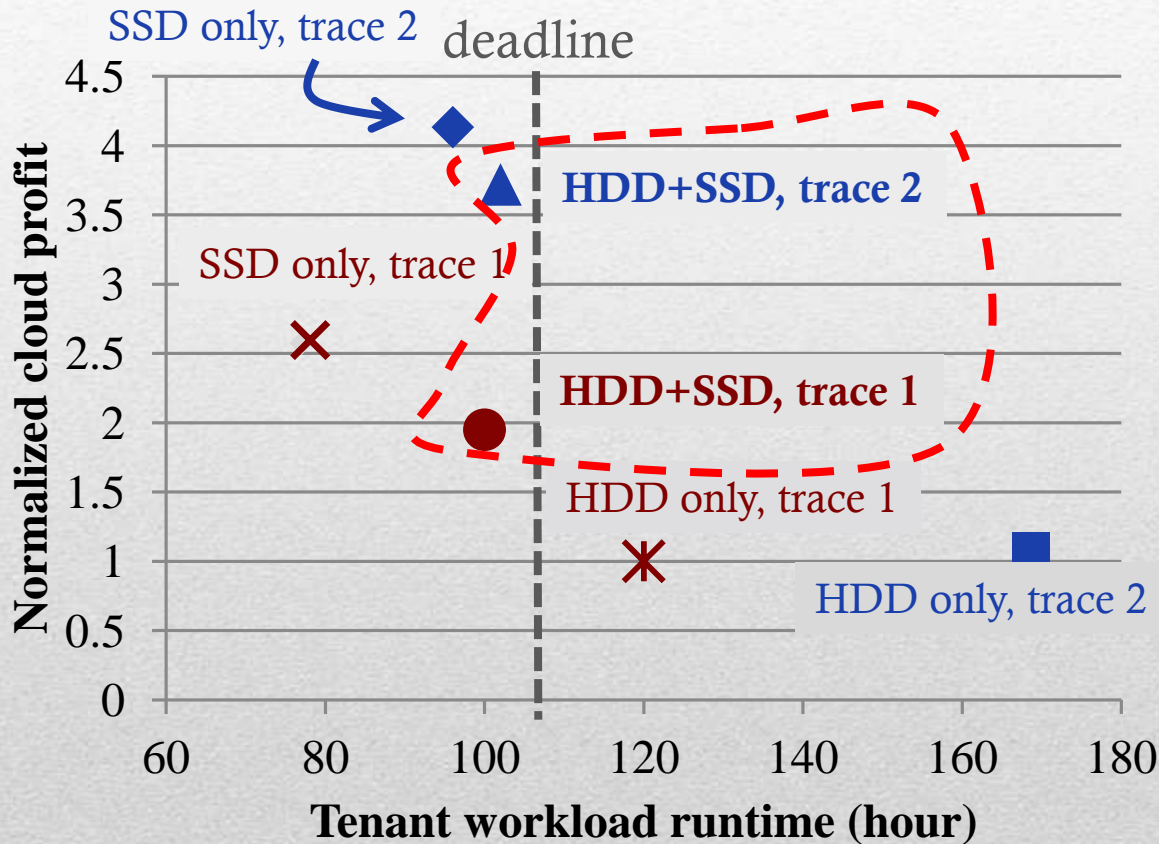
SSD-based object store helps meet workload deadline for tenants while increasing profit for provider

# SSD-based object store?





# A hybrid object store?



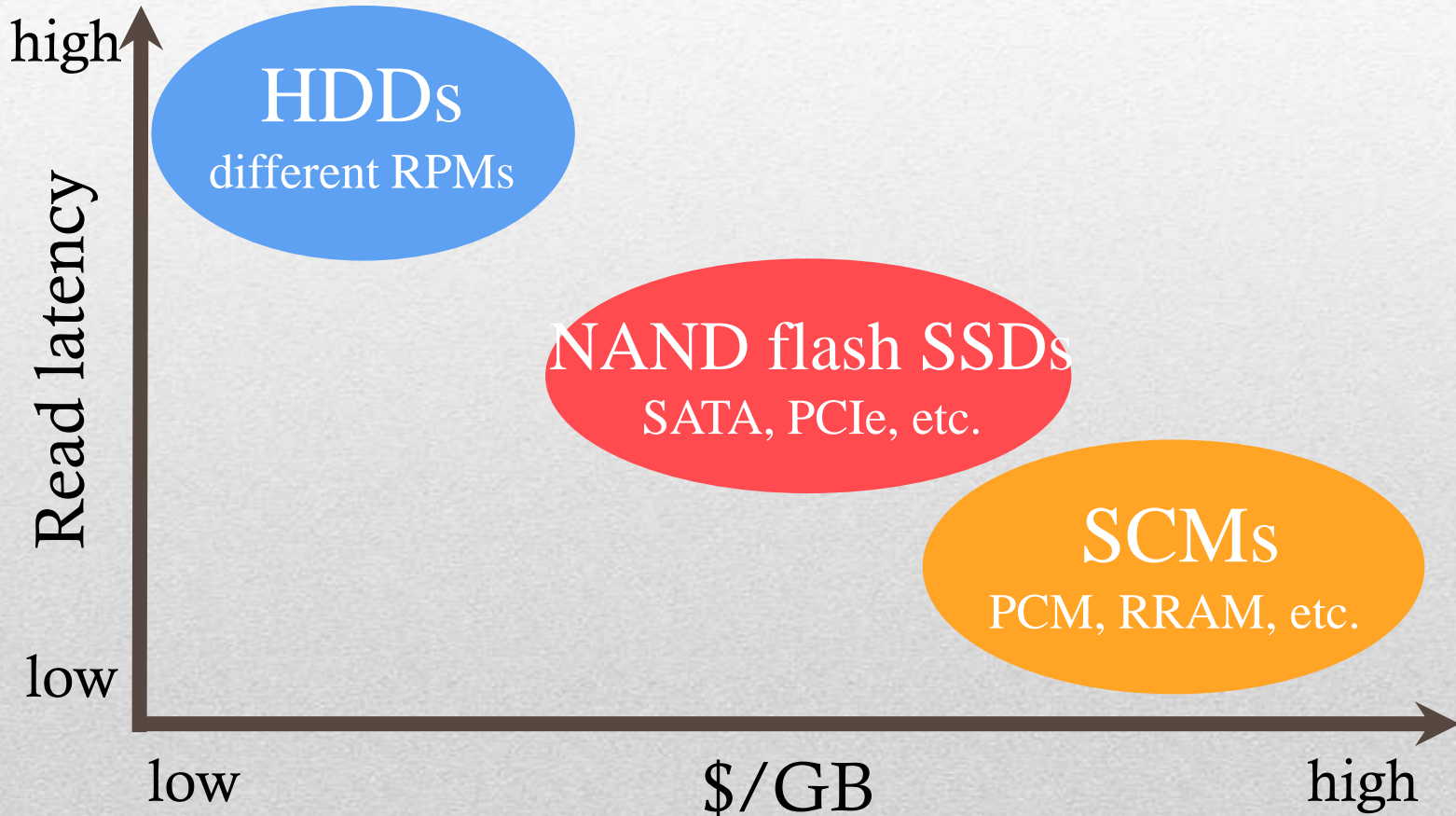
HDD price: \$0.0011/GB/day  
SSD price: \$0.0044/GB/day

Trace 1: 12TB input 5TB output  
Trace 2: 18TB input 8TB output

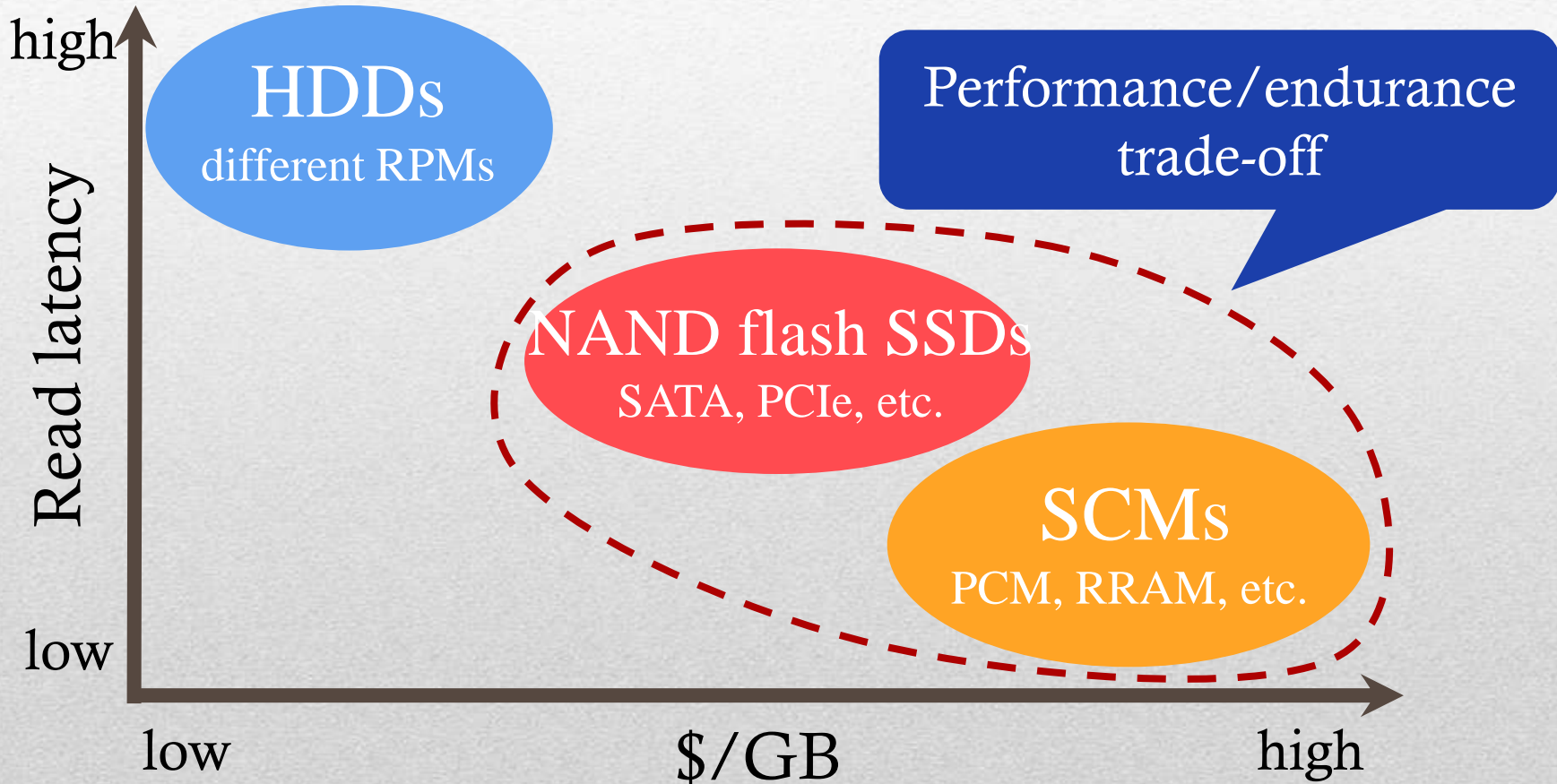
😊 Tenant: workloads meet deadline

😊 Provider: comparatively high profit

# More options in a hybrid setup

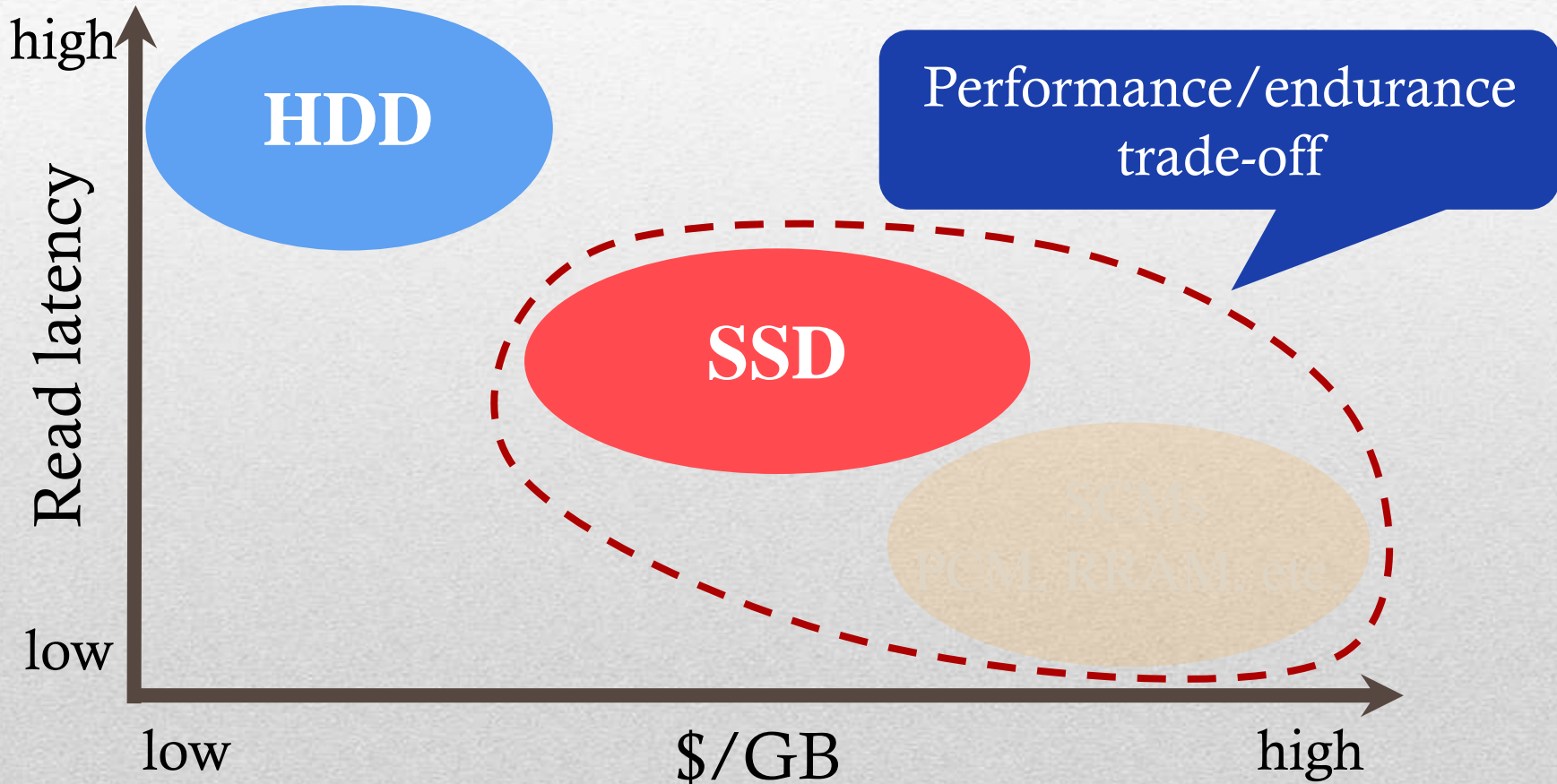


# More options in a hybrid setup





# More options in a hybrid setup



# A dynamic pricing model

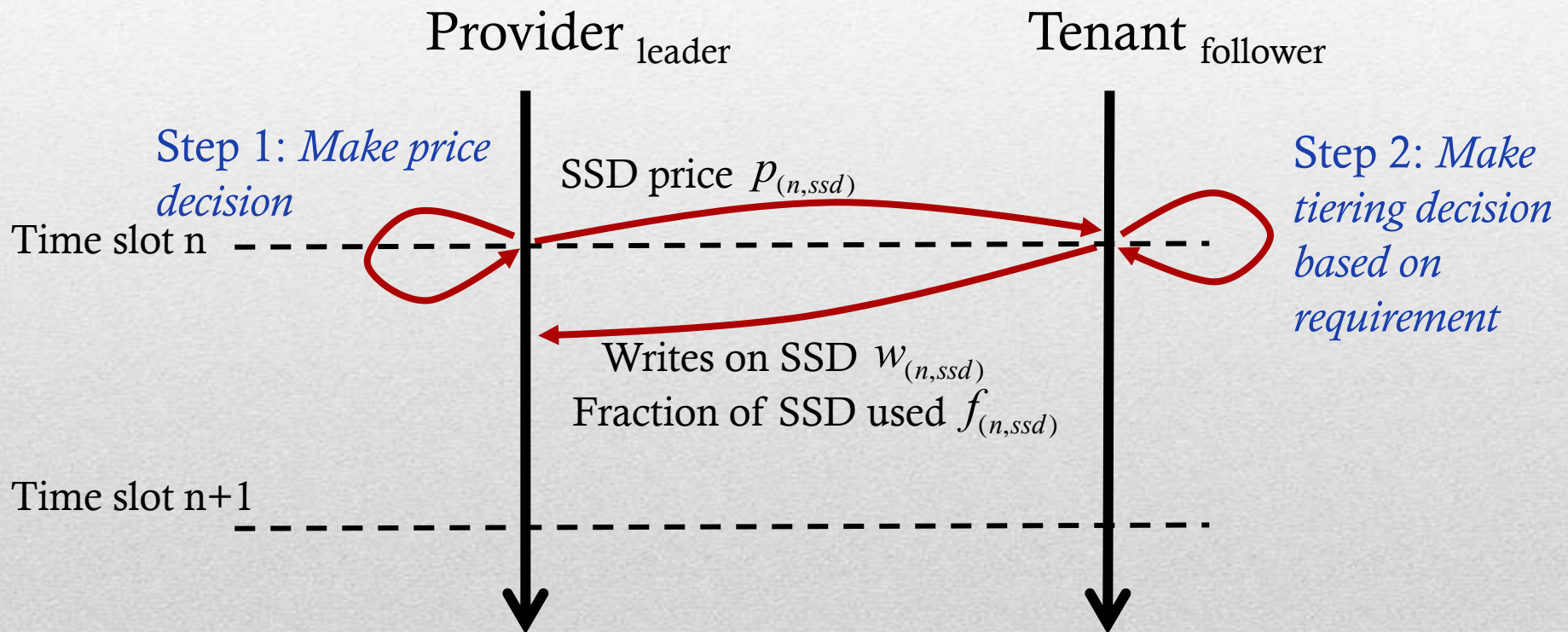
- Two objectives
  - Objective 1: to balance the increasing profit and SSD wear-out rate
  - Objective 2: to provide incentivizing mechanism to tenants

# A dynamic pricing model

- Dynamic pricing engages **both provider and tenants** in a **pricing game**
- Objectives of provider and tenants are essentially conflicting!



# The leader/follower game



# Provider model

SSD writes

% SSD demand

to maximize

$$Profit = \sum_i \left( \sum_s (Capacity_{(i,s)} \cdot f_{(i,s)} \cdot p_{(i,s)}) - Cost_i \right)$$

SSD price as decision variable

The diagram illustrates the provider model's profit function. It shows the equation: Profit = sum over i of (sum over s of (Capacity\_{(i,s)} \* f\_{(i,s)} \* p\_{(i,s)}) - Cost\_i). Annotations include: 'SSD writes' with a red arrow pointing to 'f\_{(i,s)}'; '% SSD demand' with a red arrow pointing to 'p\_{(i,s)}'; and 'SSD price as decision variable' with a red arrow pointing to 'p\_{(i,s)}'. The text 'to maximize' is placed to the left of the equation.

# Tenant model

to maximize  $Utility = \frac{1/T}{\$}$

Simulated  
annealing



$J_0: \langle s_0, c_0 \rangle$

$J_1: \langle s_1, c_1 \rangle$

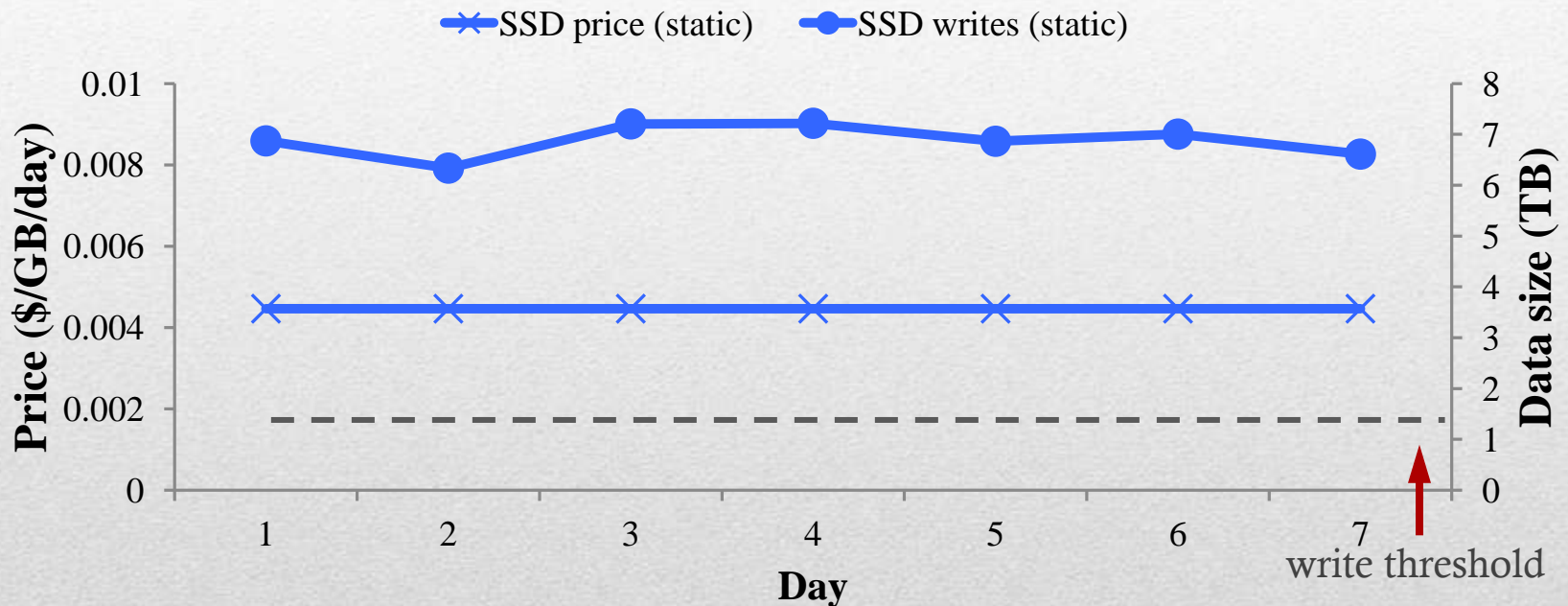
$J_2: \langle s_2, c_2 \rangle$

⋮  
Assigned job storage,  
adjusted storage capacity

\* CAST [HPDC'15]



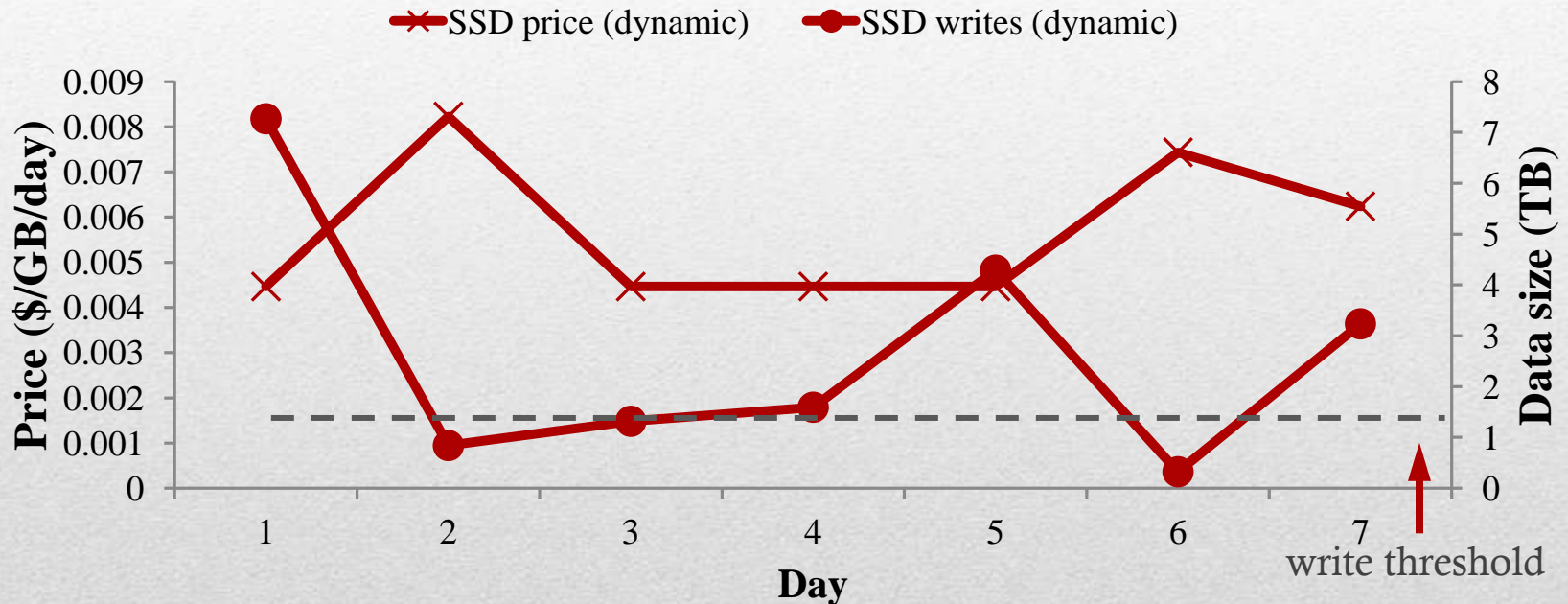
# Provider-tenant interaction



A trace snippet (250-jobs) from production trace collected from a 3000-machine Hadoop deployment at Facebook:

*The same single-day trace replayed everyday*

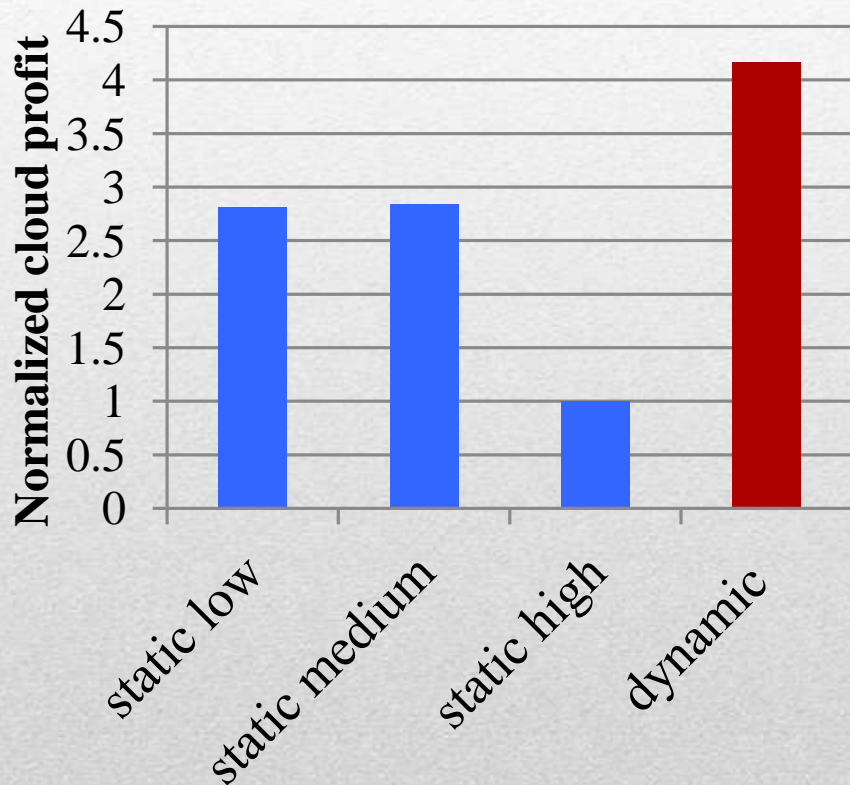
# Provider-tenant interaction



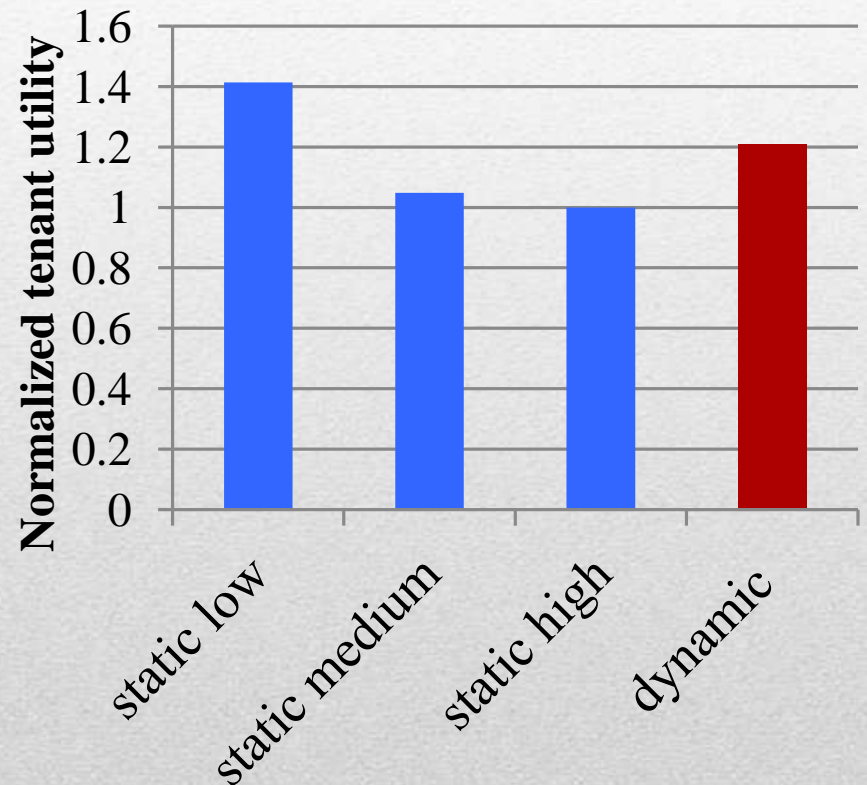
A trace snippet (250-jobs) from production trace collected from a 3000-machine Hadoop deployment at Facebook:

*The same single-day trace replayed everyday*

# Impact of different SSD pricing



static low: \$0.0035/GB/day  
static high: \$0.0121/GB/day



static medium: \$0.0082/GB/day



# Summary

Dynamic pricing + object storage tiering =  
cloud profit ↗ + tenant utility ↗