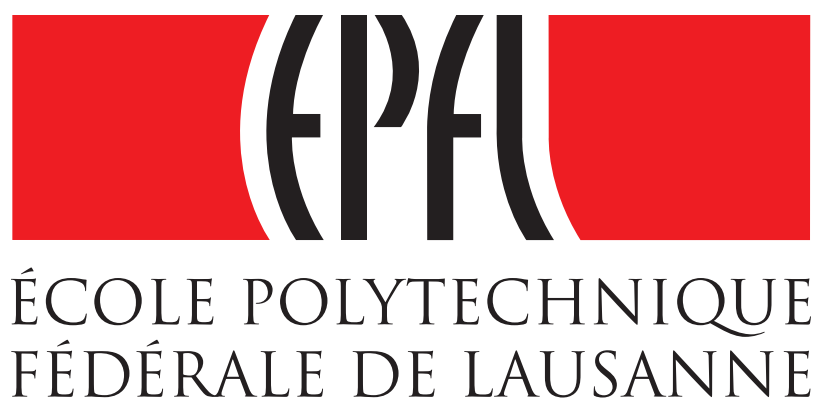
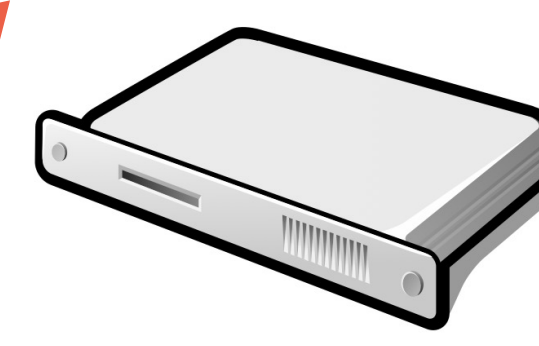
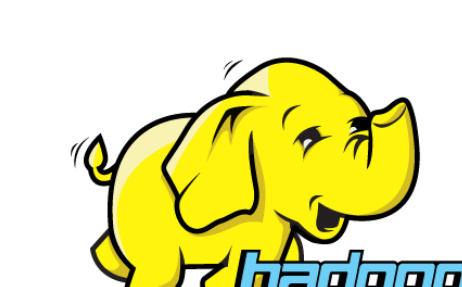



# History-Based Harvesting of Spare Cycles and Storage in Large-Scale Datacenters

Yunqi Zhang, George Prekas, Giovanni M. Fumarola, Marcus Fontoura, Íñigo Goiri, Ricardo Bianchini



## Goal: Improve Datacenter Efficiency

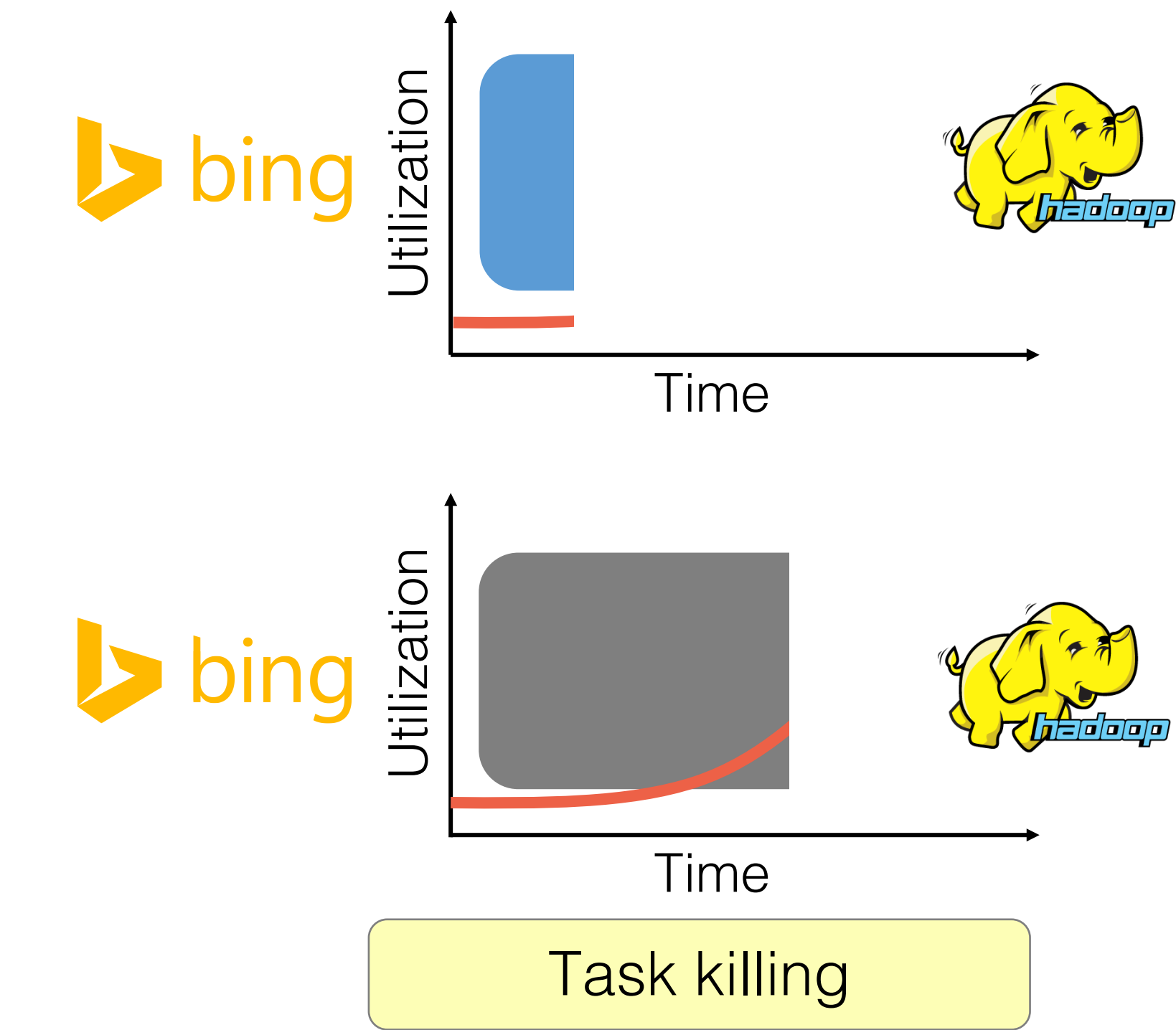


- Interactive services own the servers
- Highly over-provisioned resources for low latency
- Spare resources can be harvested
- Cannot affect interactive services performance

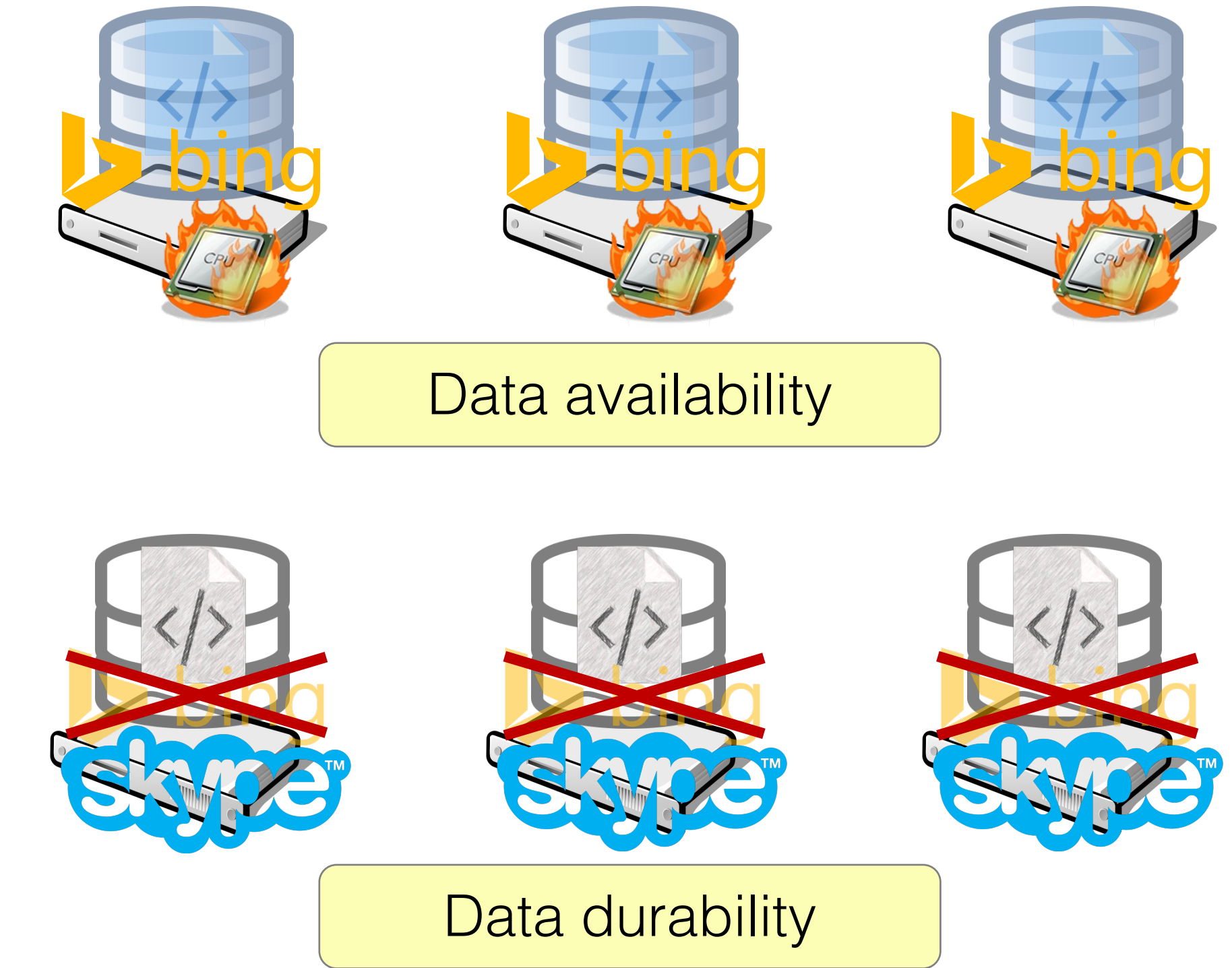
Co-locate online interactive services with batch jobs

## Challenges

Resource availability dynamics

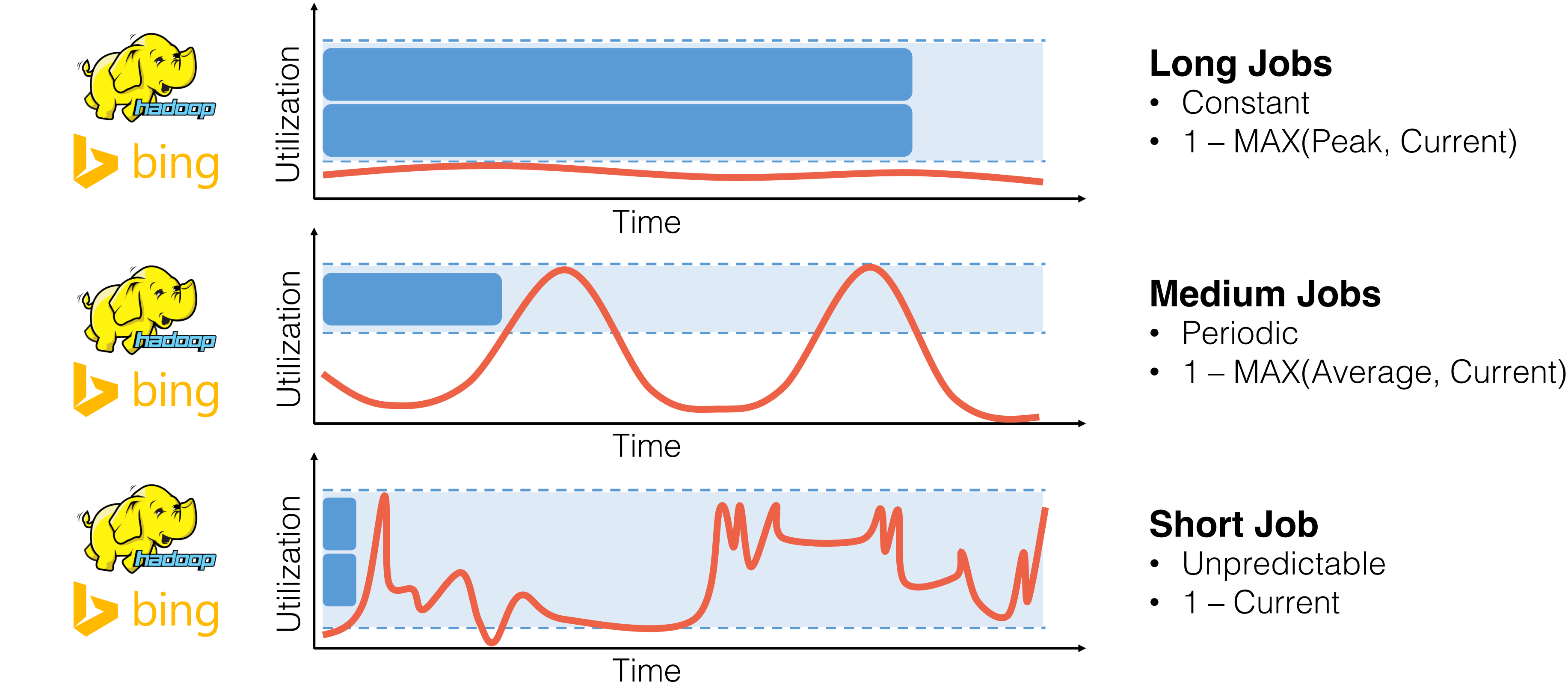


Data storage co-location

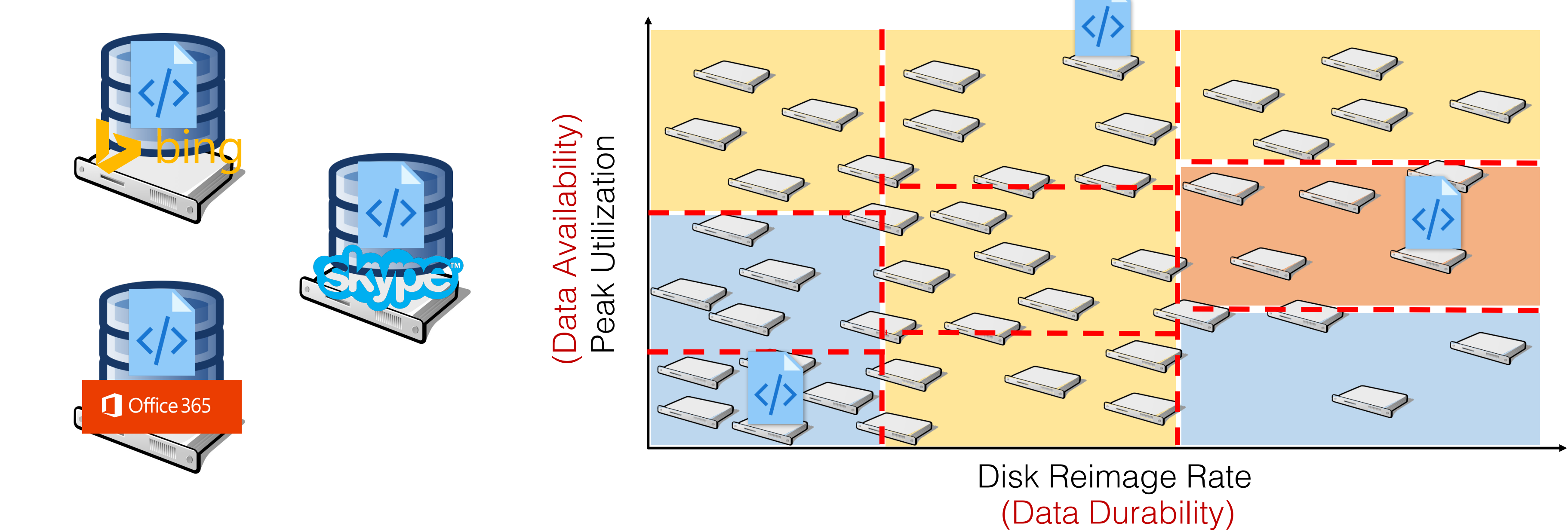


## Solution: History-Based Harvesting

History-based batch task scheduling



History-based replica placement

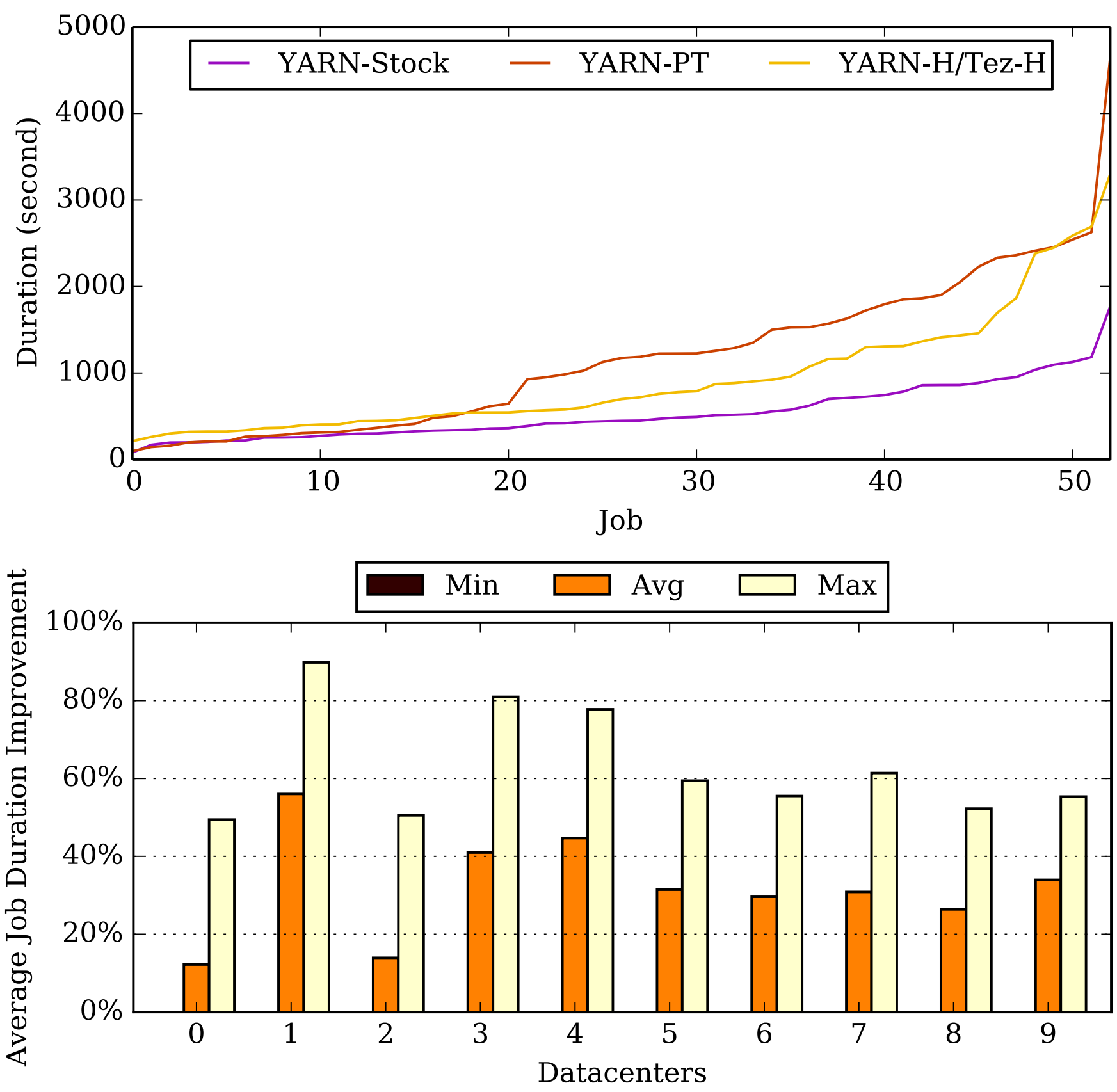


## System Implementation

System	Major Extension
YARN-H	Report primary tenant utilization to the RM Kill containers due to primary tenant needs
Tez-H	Leverage information on the observed job lengths Schedule tasks on servers unlikely to kill them
HDFS-H	Track primary tenant utilization, deny accesses Place replicas at servers with diverse patterns

## Evaluation

Batch task scheduling



Data replica placement

