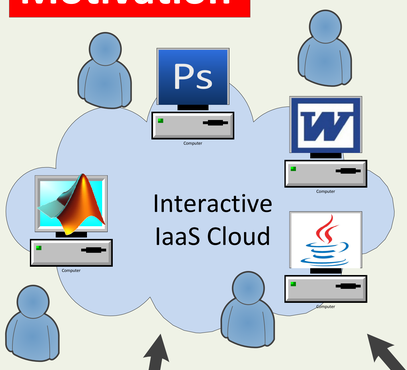
# Augmenting MapReduce with Active Volunteer Resources

R Benjamin Clay, Zhiming Shen, Xiaosong Ma, Xiaohui Gu

Department of Computer Science, North Carolina State University, Raleigh, North Carolina, USA

NC STATE UNIVERSITY

## Motivation

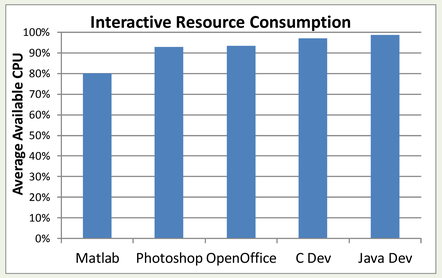



### Story Overview

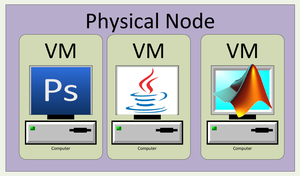
1. Significant residual resources exist in interactive IaaS clouds

2. Padding with a batch workload is more flexible than consolidation

3. Dynamic cloud conditions necessitate active management




### Traditional Solution: Consolidation




### Migration Ill-Suited for Interactive Workloads

| | Matlab | Photoshop | OpenOffice | C Dev |
|---|---|---|---|---|
| Burst Height | 39.9 % | 25.8 % | 31.0 % | 27.7 % |
| Burst Length | 6.9 sec | 2.0 sec | 1.3 sec | 47.4 sec |
| Offline Migration | 5.7 min | | | |

Short bursts relative to migration

| | | | | |
|---|---|---|---|---|
| Reservation Avg | 92.5 min | 74.0 min | 70.2 min | 120.1 min |
| Reservation StdDev | 90.3 min | 78.9 min | 90.9 min | 99.1 min |

Highly variable reservations

- Burst data collected from traces of real interactive users
- Migration data from **offline, non-shared-storage** migration
  -> 30gb VM image
  -> unloaded 1 GigE crossbar network.
- Reservation data collected by NCSU VCL, 2004-2010.

Virtual Computing Lab @ NCSU
- 800 VCL + 800 HPC blades
- 151 VM images
- 30k users
- 700k reservations since 2004
- 8 major universities in NC

Amazon EC2
- 11 node types
- 1008 VM images
- 6 regions
- Spot Instances via auctions
- ~70k instance launches / day

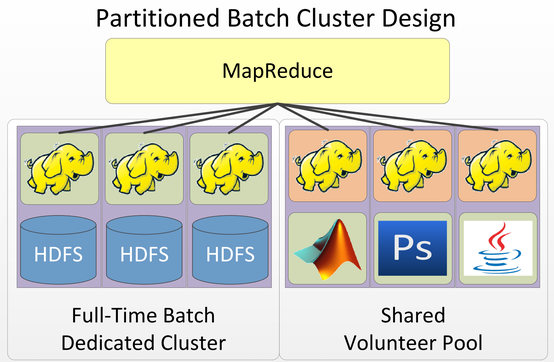
## Solution Overview

### Batch Workload Can Use Idle Cycles

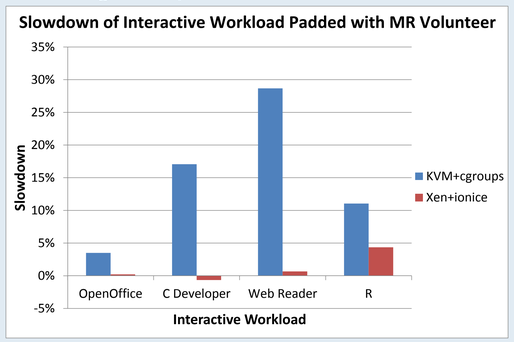
- Pad interactive VMs with **batch volunteers**
  -> Interactive demands always satisfied
  -> Eliminates wasted cycles
  -> Priority enforced using hypervisor
- Batch workloads ideally suited for environment
  -> Throughput oriented
  -> Latency-insensitive = burst-tolerant
- Supplement dedicated batch cluster with volunteers
  -> Volunteers alone too unstable
  -> Dedicated provide stability, performance baseline
  -> Dedicated provide persistent data storage

**2 tiers can capture cycles sustainably**

### Proposed 2-Tiered Solution




### Padding: Acceptable Performance Isolation



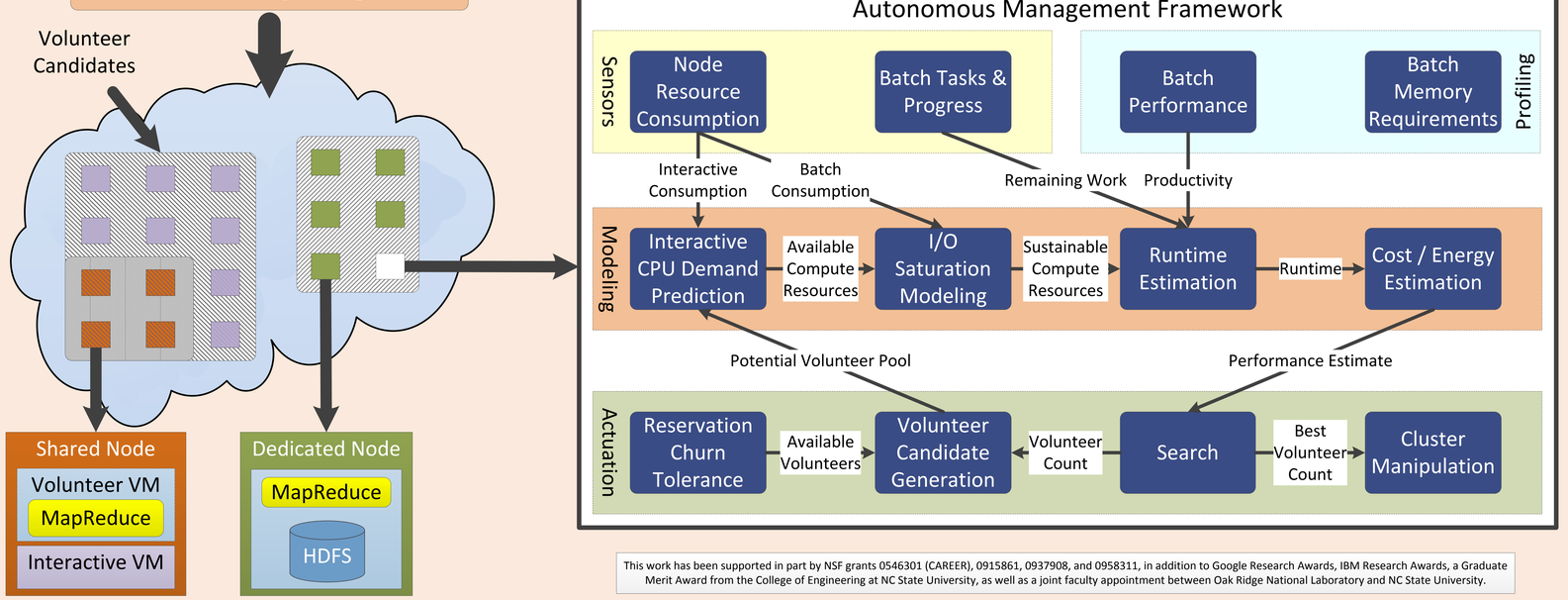

Hypervisor-enforced isolation is feasible, depending on load.
- Maximum priority given to interactive VM
- Interactive workloads from Linux bltk + AT&T R benchmark
- Tested alongside Word Co-occurrence MapReduce workload
  -> High CPU utilization, large intermediate data volume

## Management Framework

### Ideal volunteer pool size?

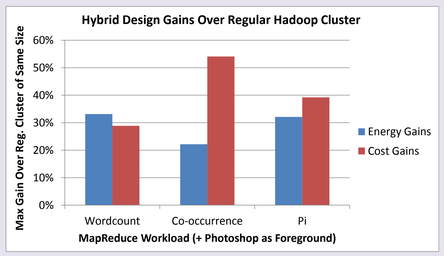
Range of contributing factors:
- Dynamic interactive demand
- Batch workload
- Dedicated cluster
- Hardware (disk, CPU, network)

**Ideal continuously changes**

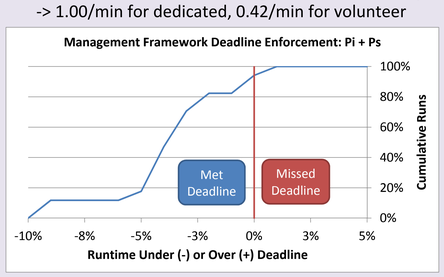


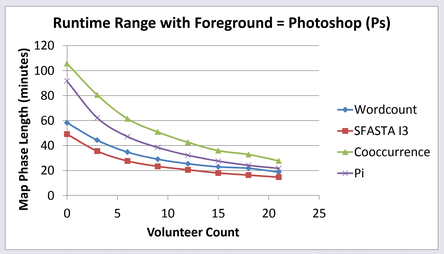

This work has been supported in part by NSF grants 0546301 (CAREER), 0915861, 0937908, and 0958311, in addition to Google Research Awards, IBM Research Awards, a Graduate Merit Award from the College of Engineering at NC State University, as well as a joint faculty appointment between Oak Ridge National Laboratory and NC State University.

## Preliminary Results




- NCSU's HGCC cluster, 4-10 nodes (Hybrid has 2 dedicated)
- Volunteers pay incremental energy above foreground
- Cost ratio based on EC2 m2.xlarge On-Demand and Spot
  -> 1.00/min for dedicated, 0.42/min for volunteer




- NCSU's ARC cluster, 0-21 volunteers w/ 6 dedicated
- Deadline results below achieved on same deployment
  -> Both I/O- (Wordcount) and CPU- (Pi) intensive shown
- On-going work: Automated cost and energy minimization