CASE STUDY: COST CALCULATOR FOR CLOUD APPLICATIONS

Background:

Salánki, Á., Kincses, G., Gönczy, L. and Kocsis, I., 2017. Data analysis-based capacity planning of VCL clouds. *International Journal of Cloud Computing*, *6*(4), pp.370-383.

Motivation

Enterprise cloud
Purchased CPU time

Azure

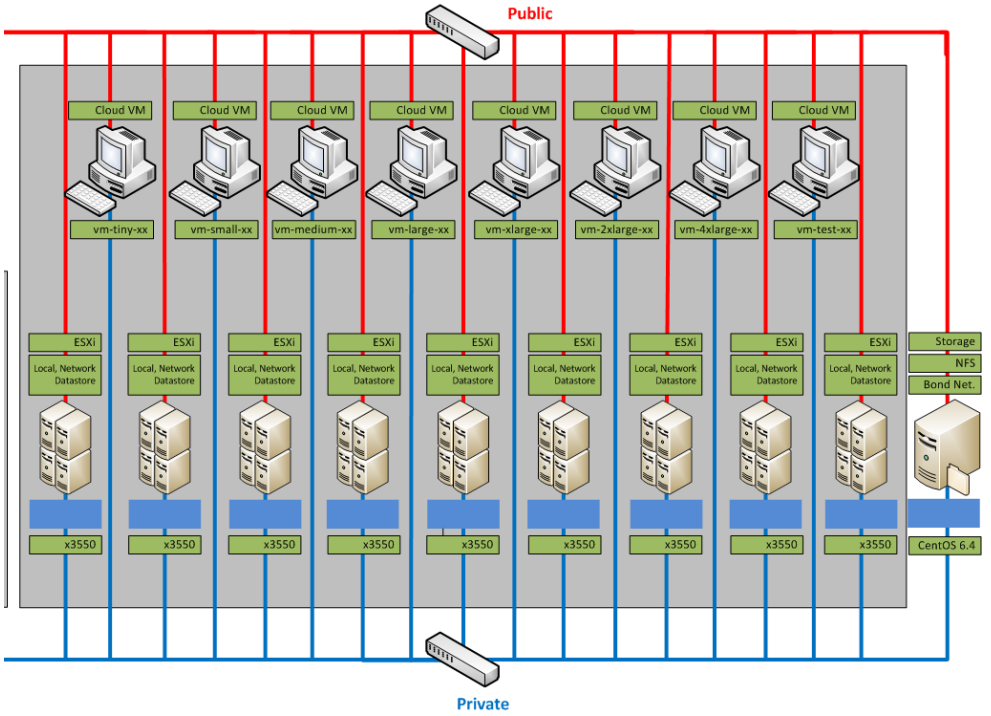
amazon web services™

Lab

Private university cloud

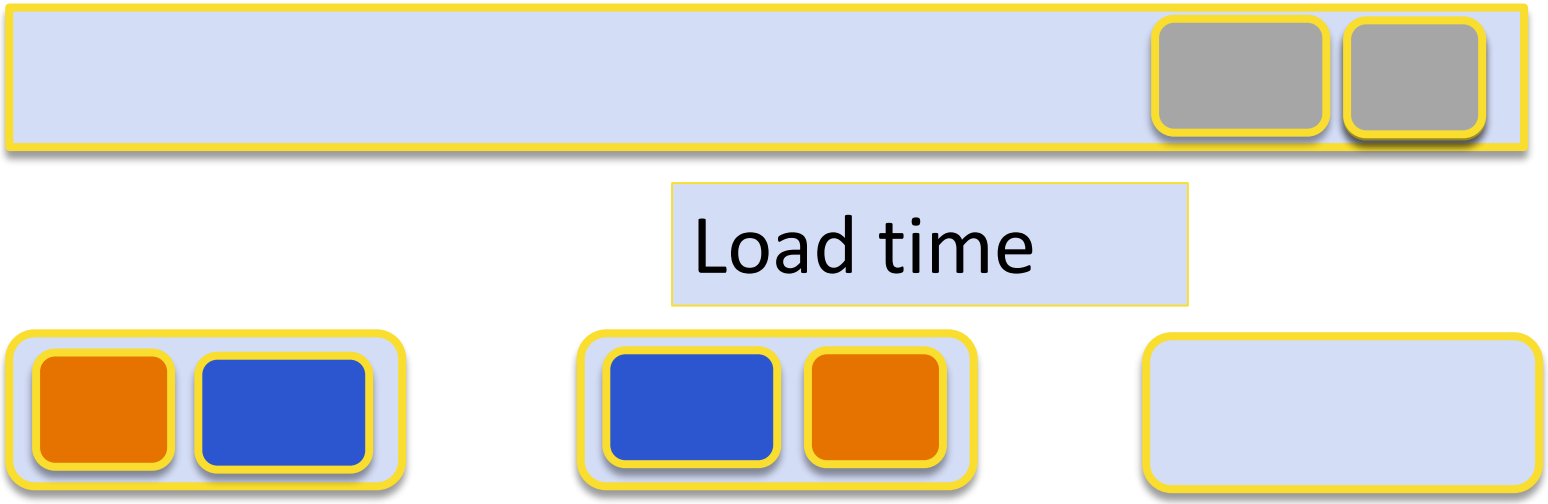
Our VCL cloud

- Maintained by our research group
- 5 semesters
  - 2 courses/semester
- 9 hosts
- ~20 000 reservations
  - Only 22 rejected

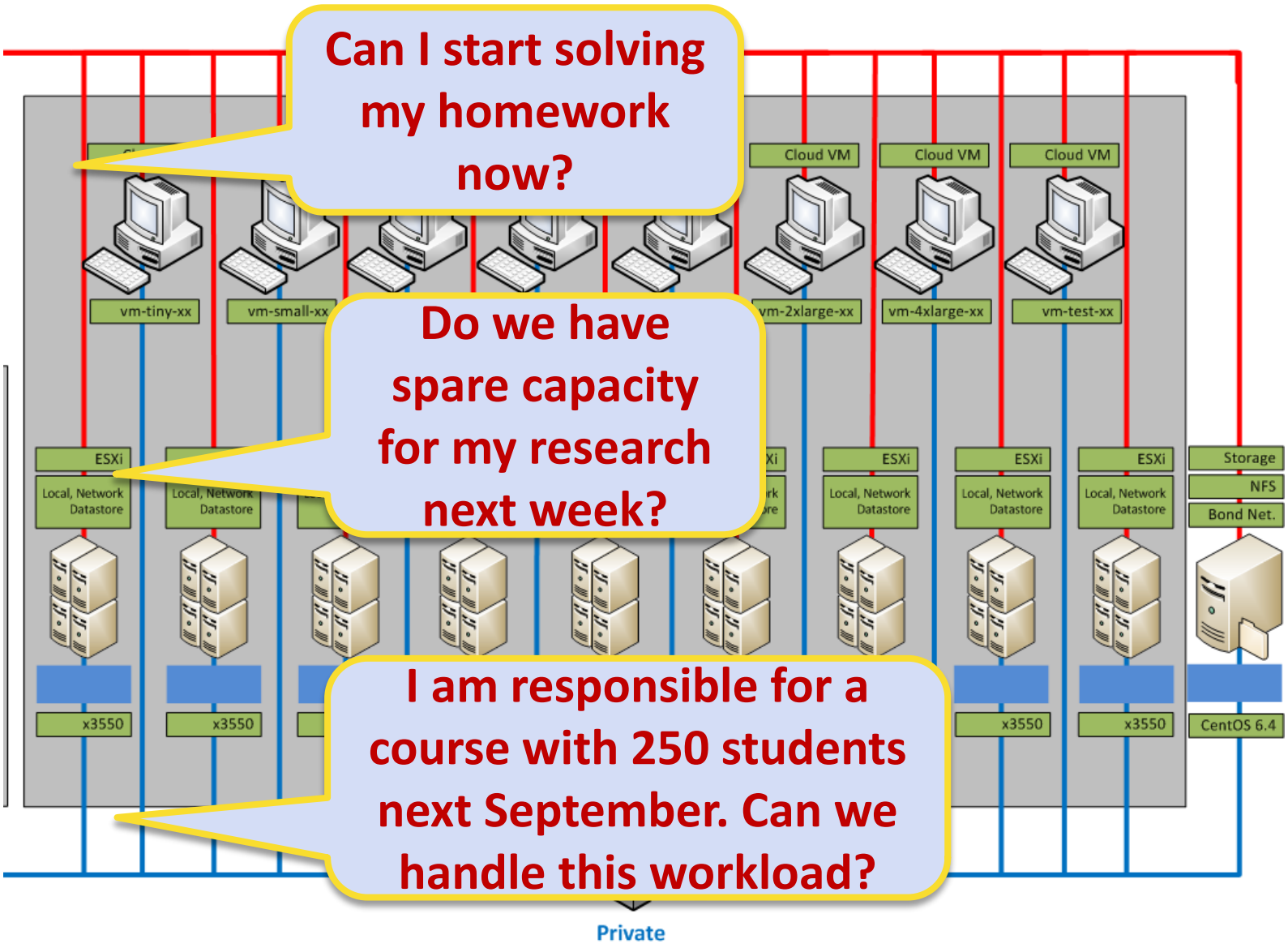


Reservation Workflow in VCL

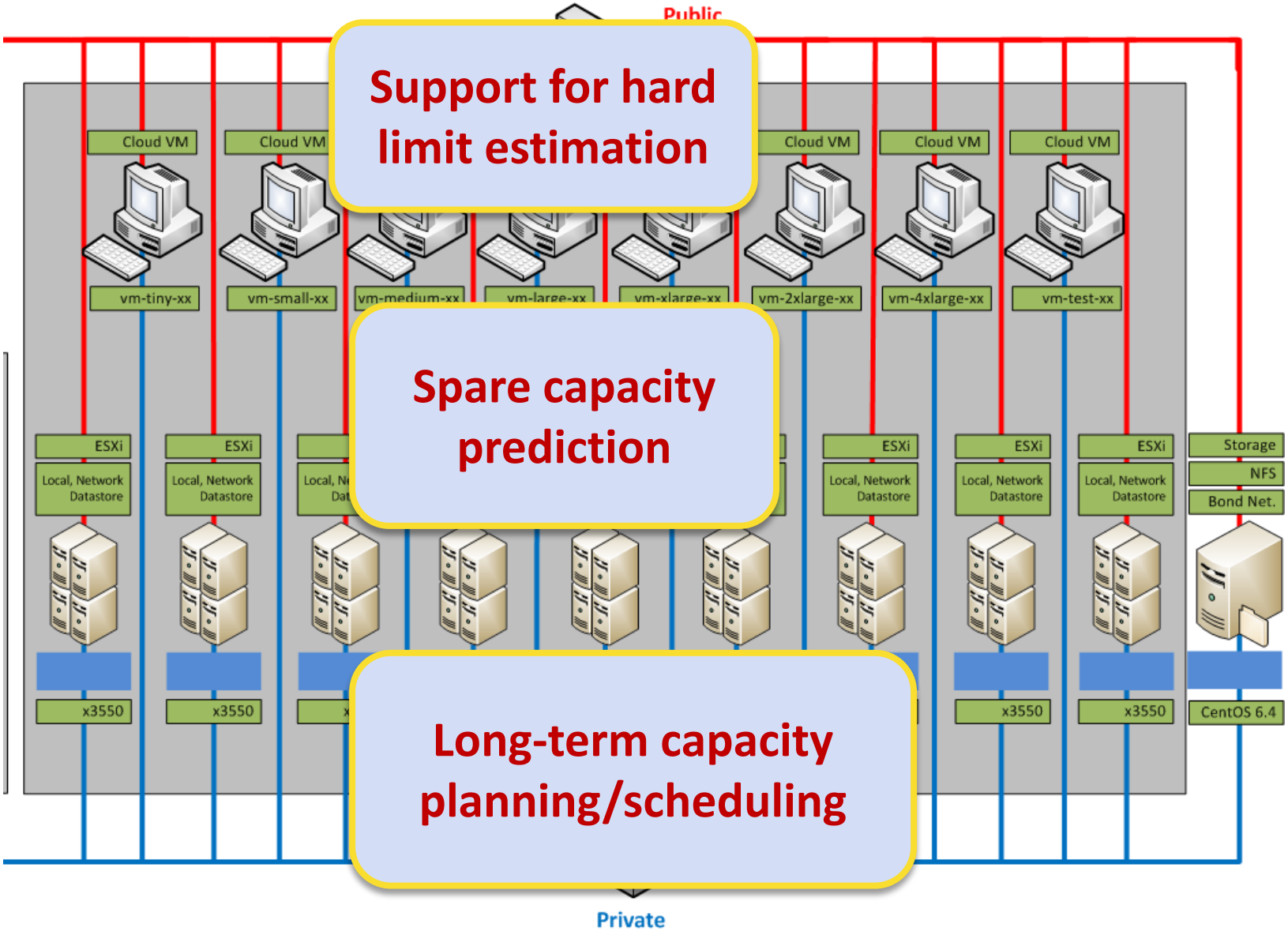
- Request
  - VM type
  - Length
  - Immediately or later
- Hard reservation limit



Capacity Planning



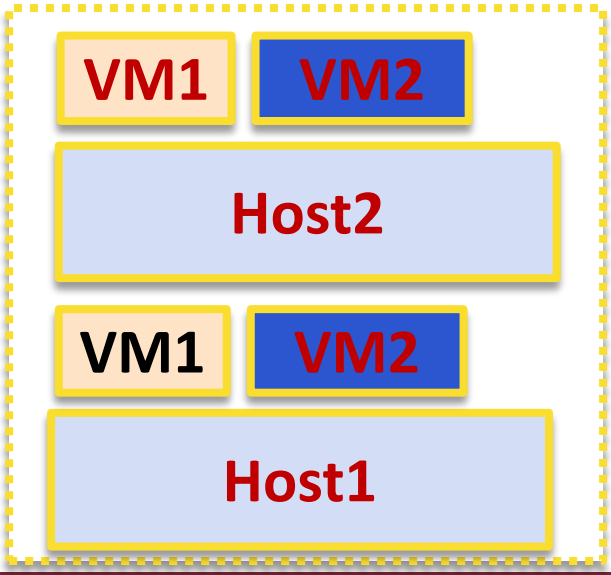
Capacity Planning



The Available Dataset

deadlines, #students

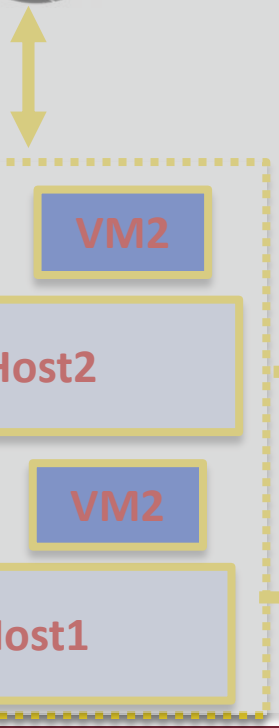
reservation type, time to load, etc.

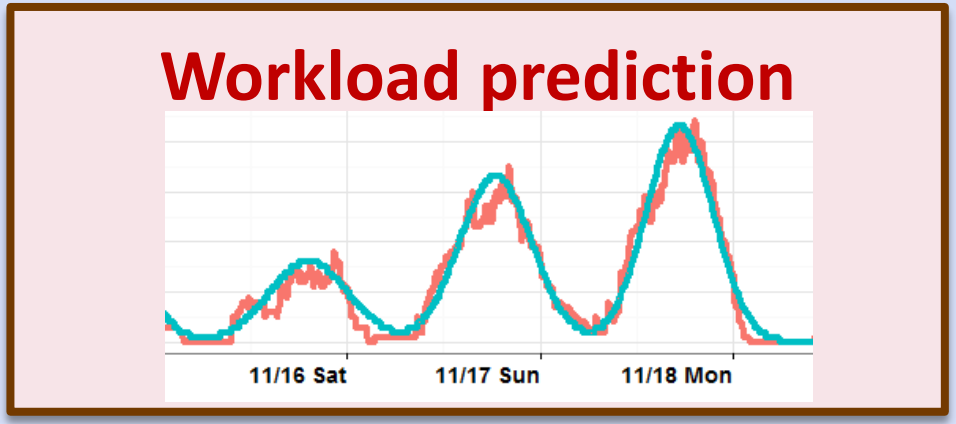


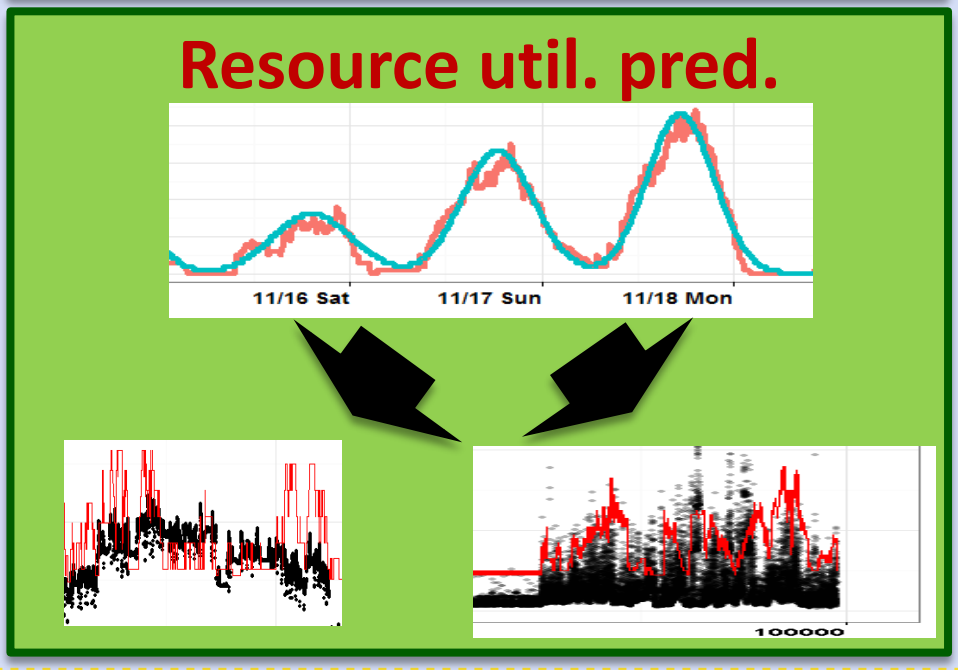
cpu usage, memory usage, etc.

cpu usage, memory usage, etc.

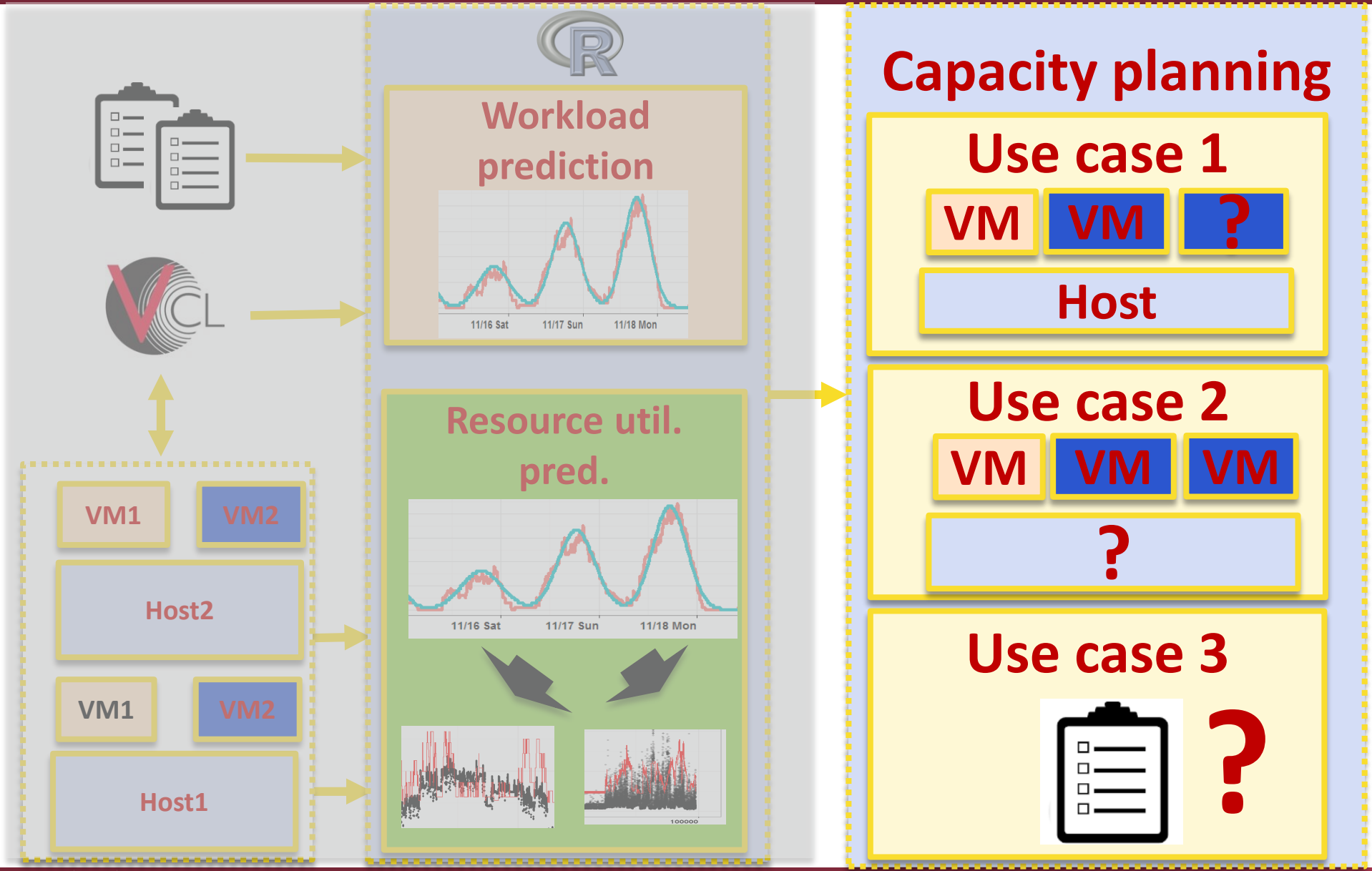
Data Analysis Steps



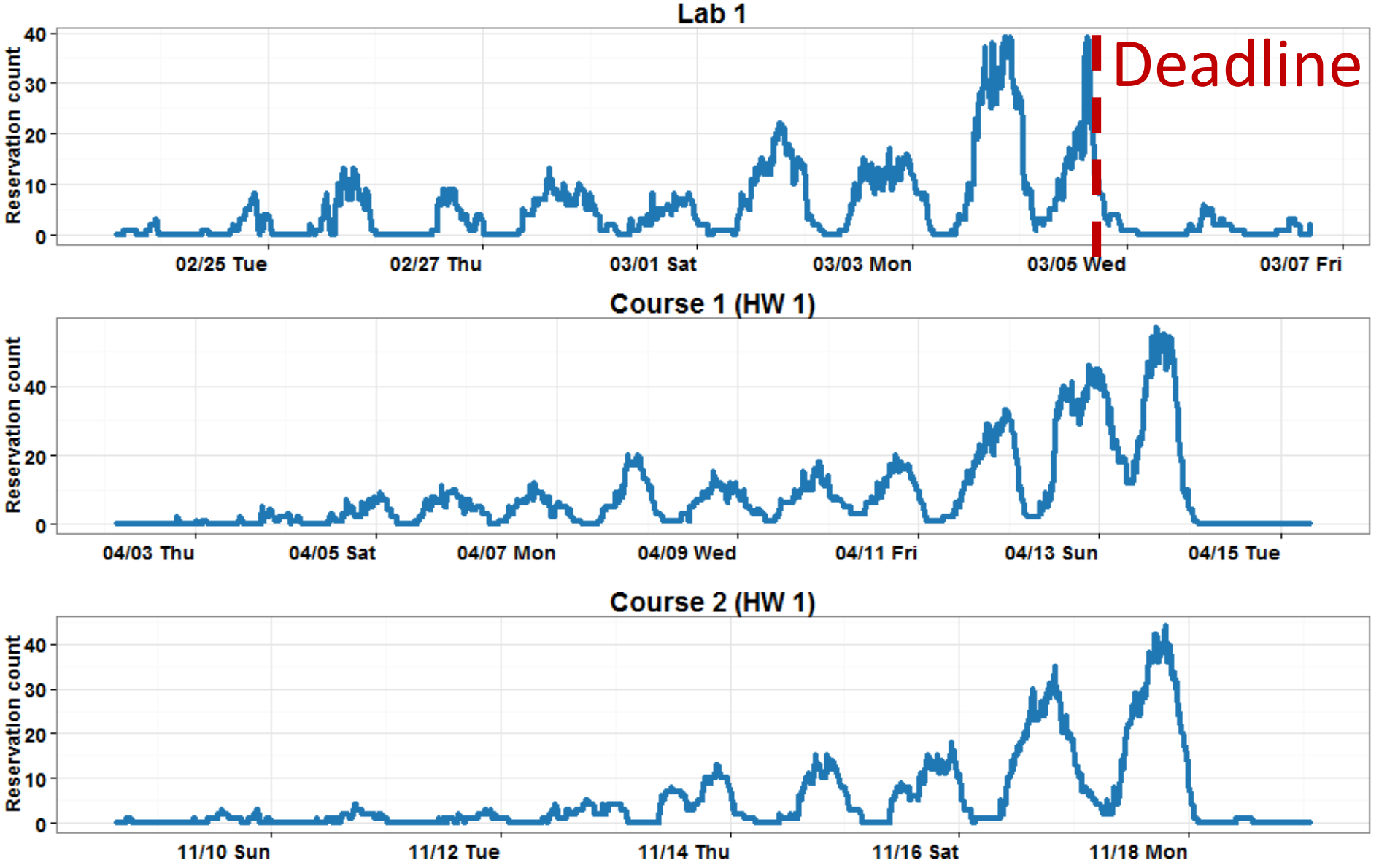




Workflow

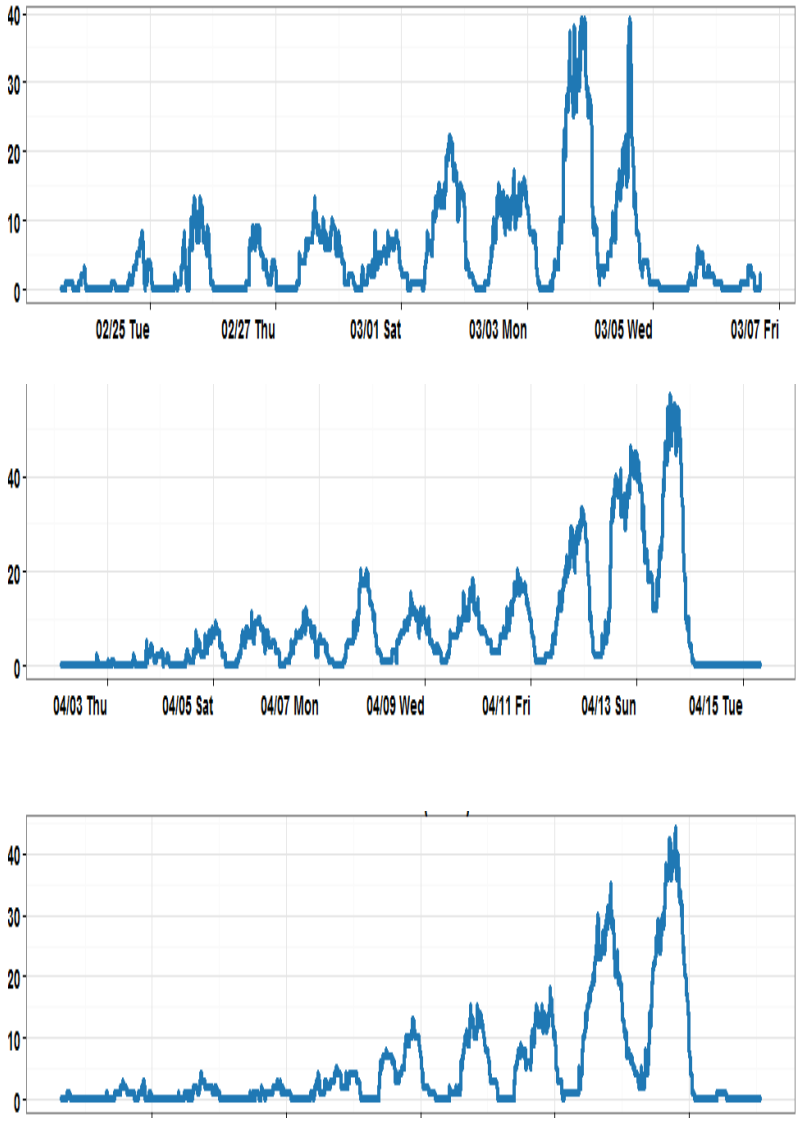

Workload prediction
11/16 Sat
11/17 Sun
11/18 Mon
Resource util. pred.
11/16 Sat
11/17 Sun
11/18 Mon
VM1
VM2
Host2
VM1
VM2
Host1
Capacity planning
Use case 1
VM
VM
?
Host
Use case 2
VM
VM
VM
?
Use case 3
?

Workload prediction

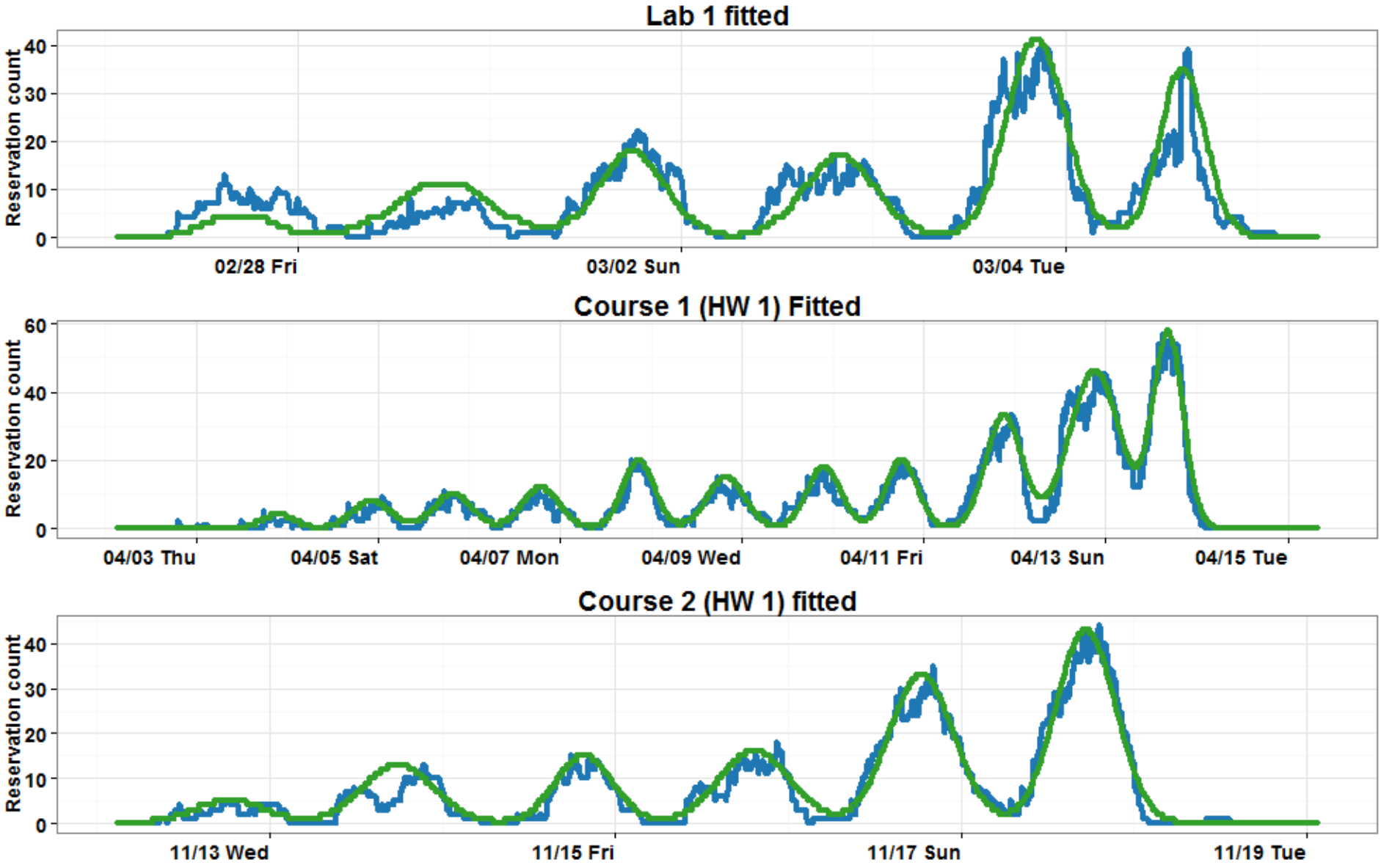


Workload prediction

- Daily workload follows a Gaussian-like distribution

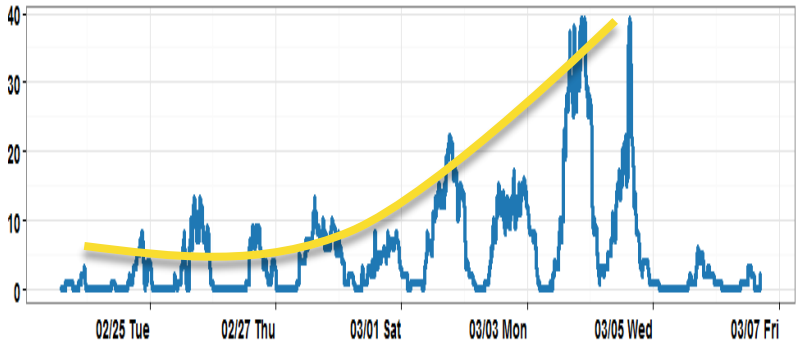


Model fitting

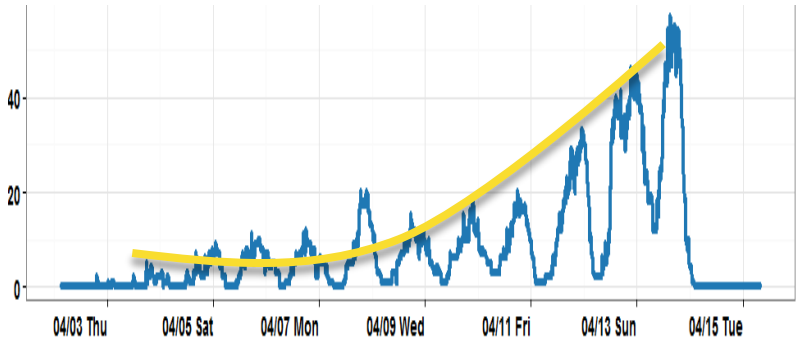


Workload prediction

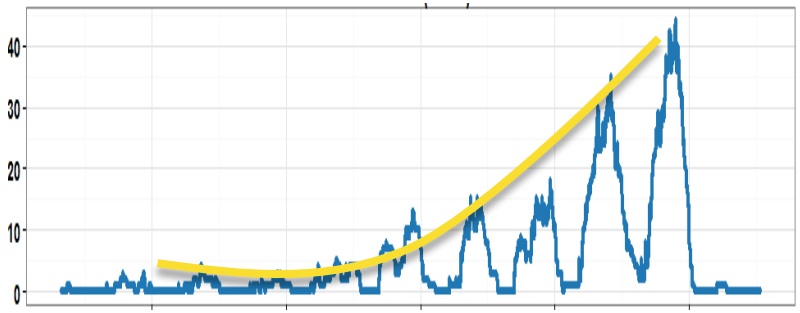
- Daily workload follows a Gaussian-like distribution



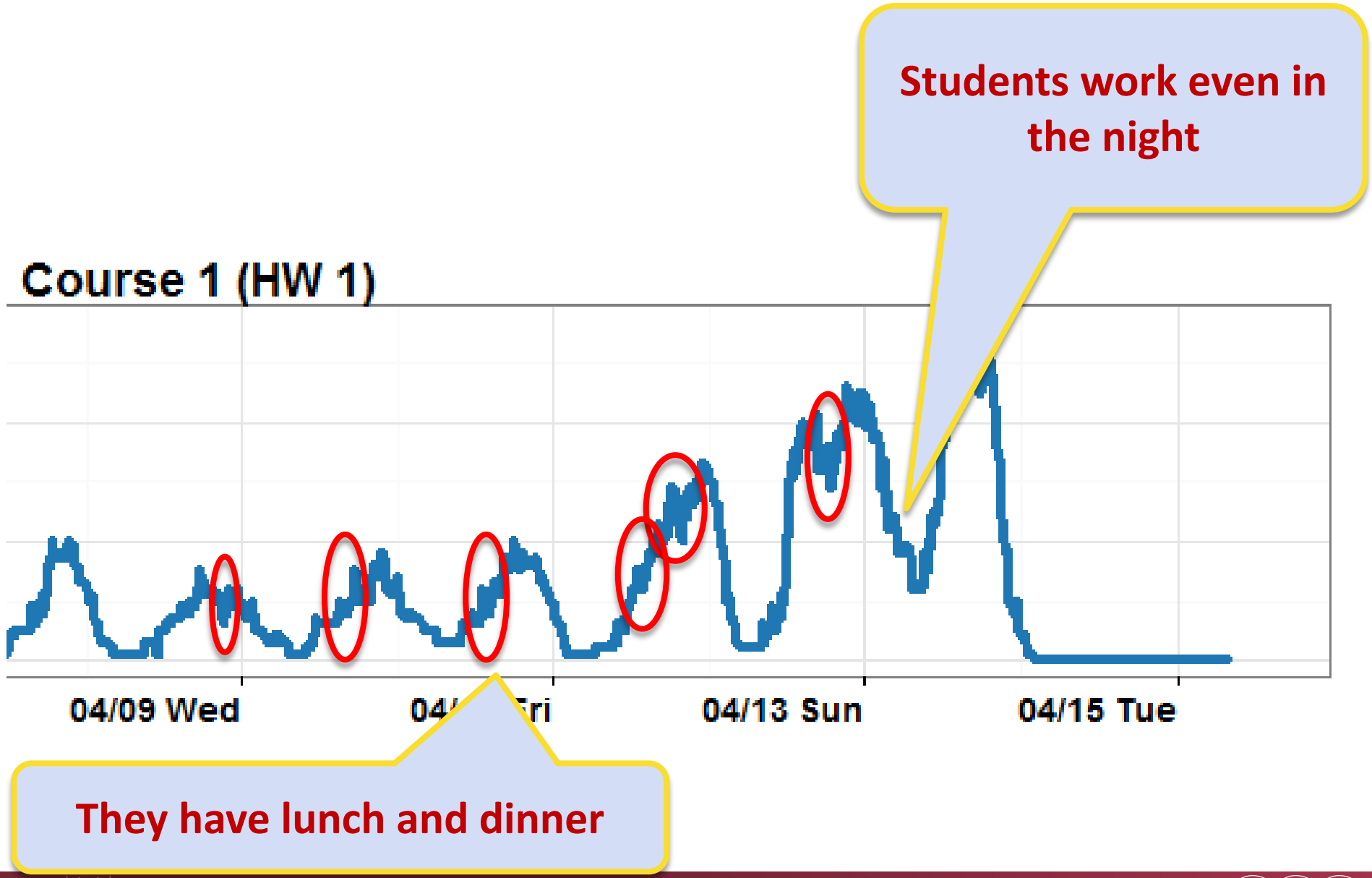
- Exponential increase in peak numbers
- maximum location between 7 PM and 11 PM



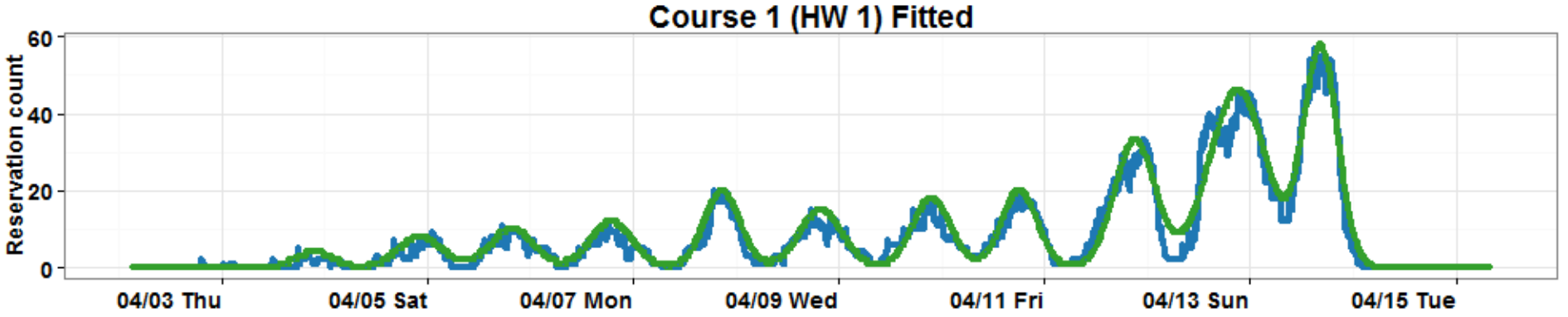
- ~4 hours as standard deviation



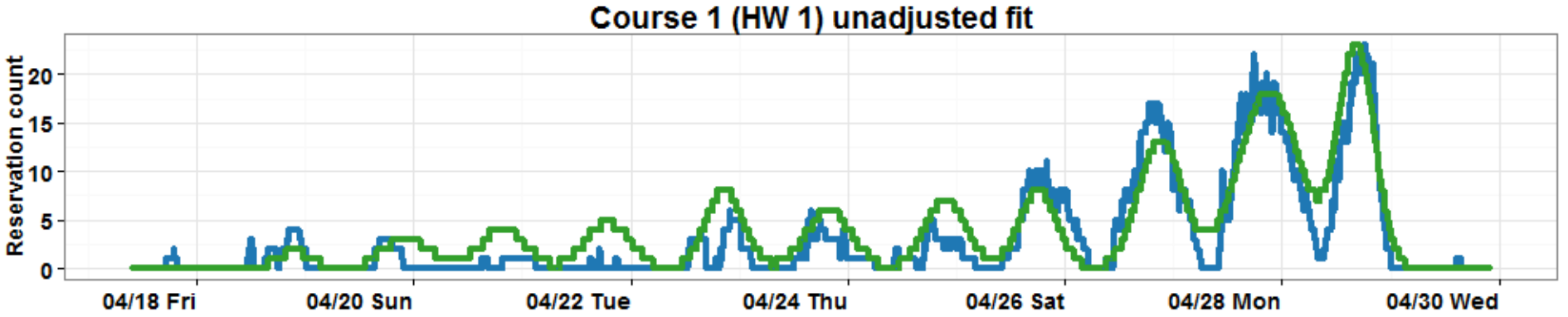
Workload prediction



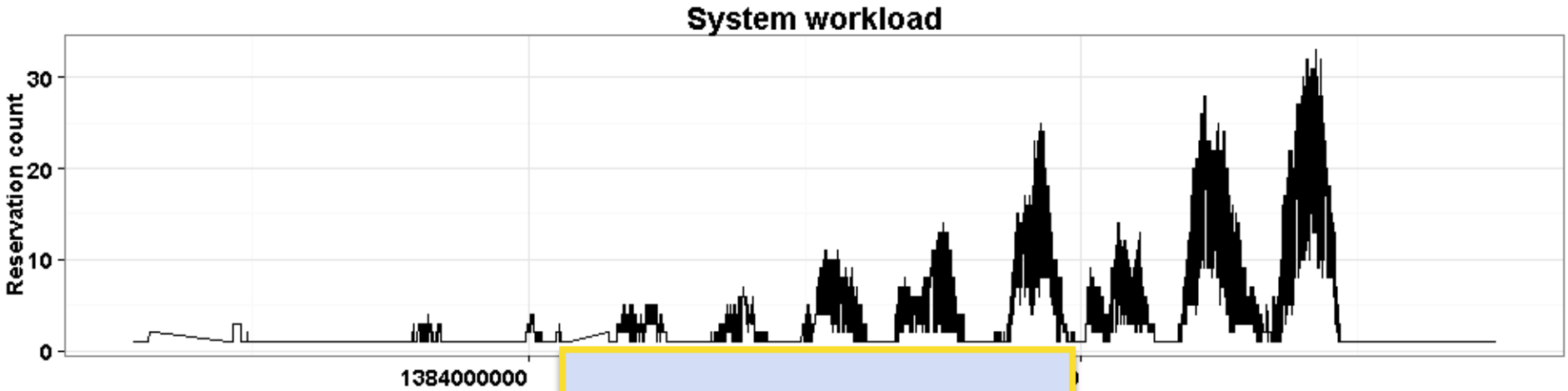
Workload prediction



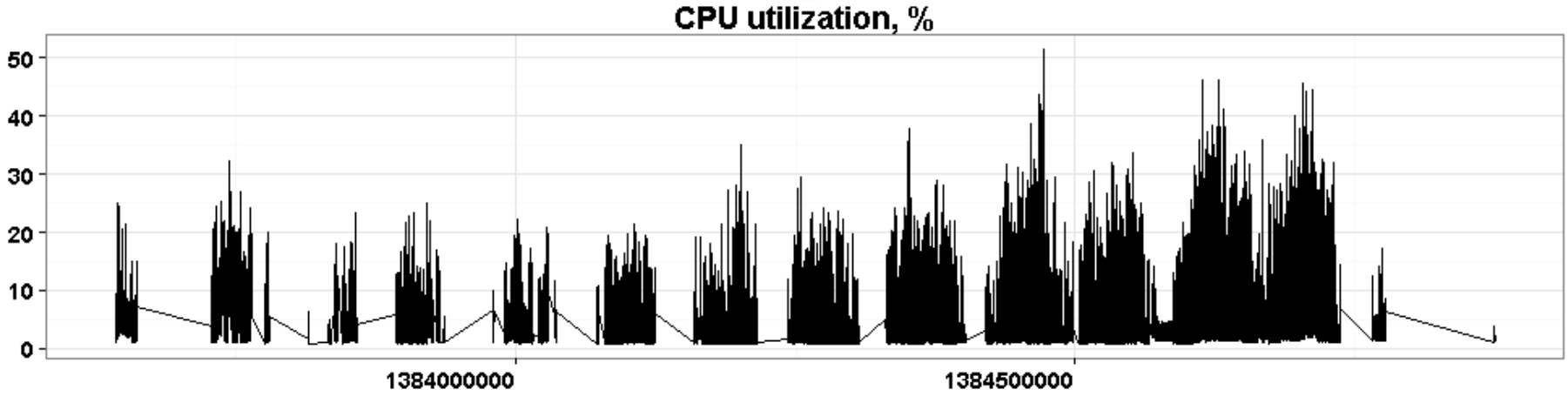
Changes in students' behavior?



Resource Utilization Prediction

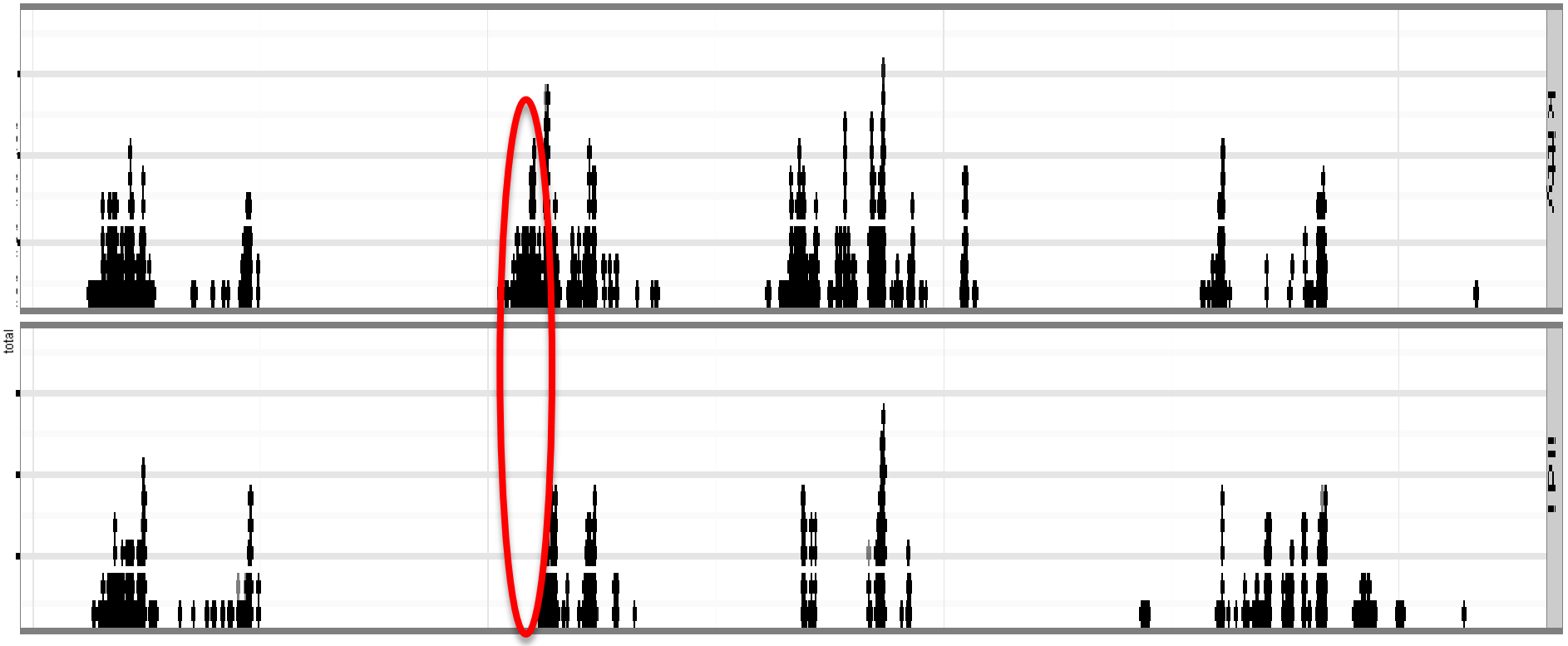


$f(workload)$??



Challenges

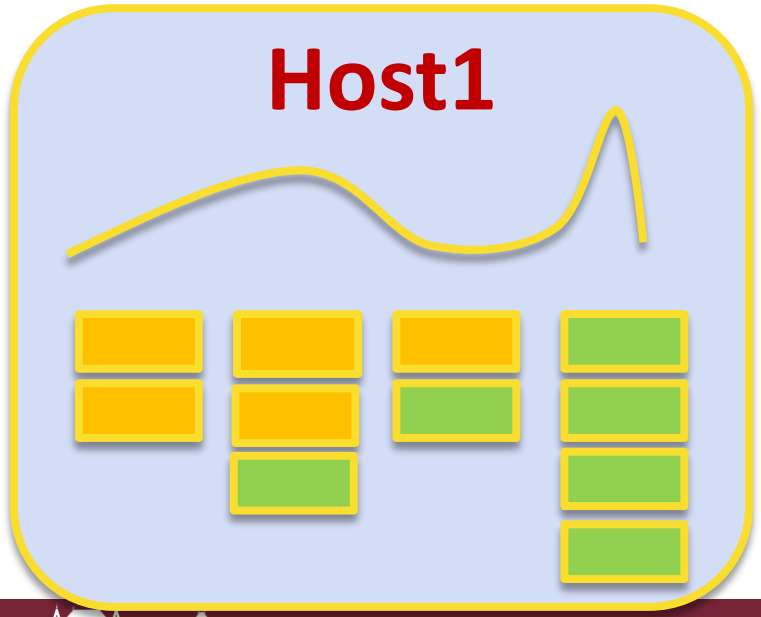
- It is a cloud
  - Statistical multiplexing ☺

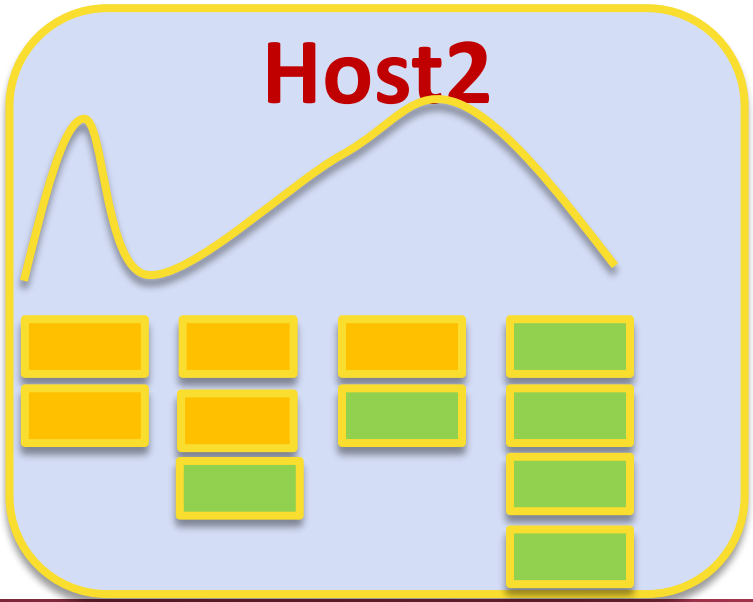


2012/2013/2 | 2013/2014/1 | 2013/2014/2 | 2014/2015/1

Challenges

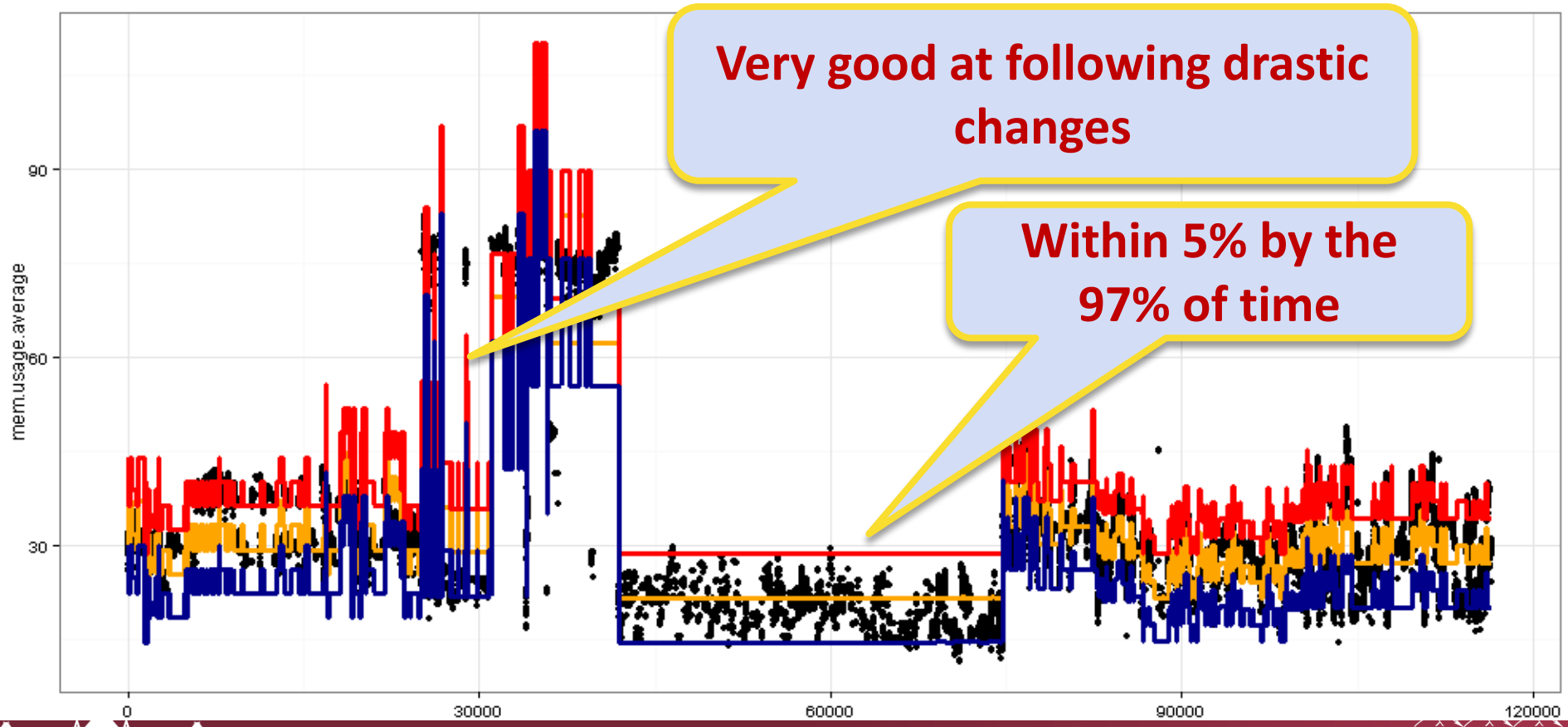
- It is a cloud
- Hosts show different behavior
  - Warm spare
  - Different user behavior
  - ???





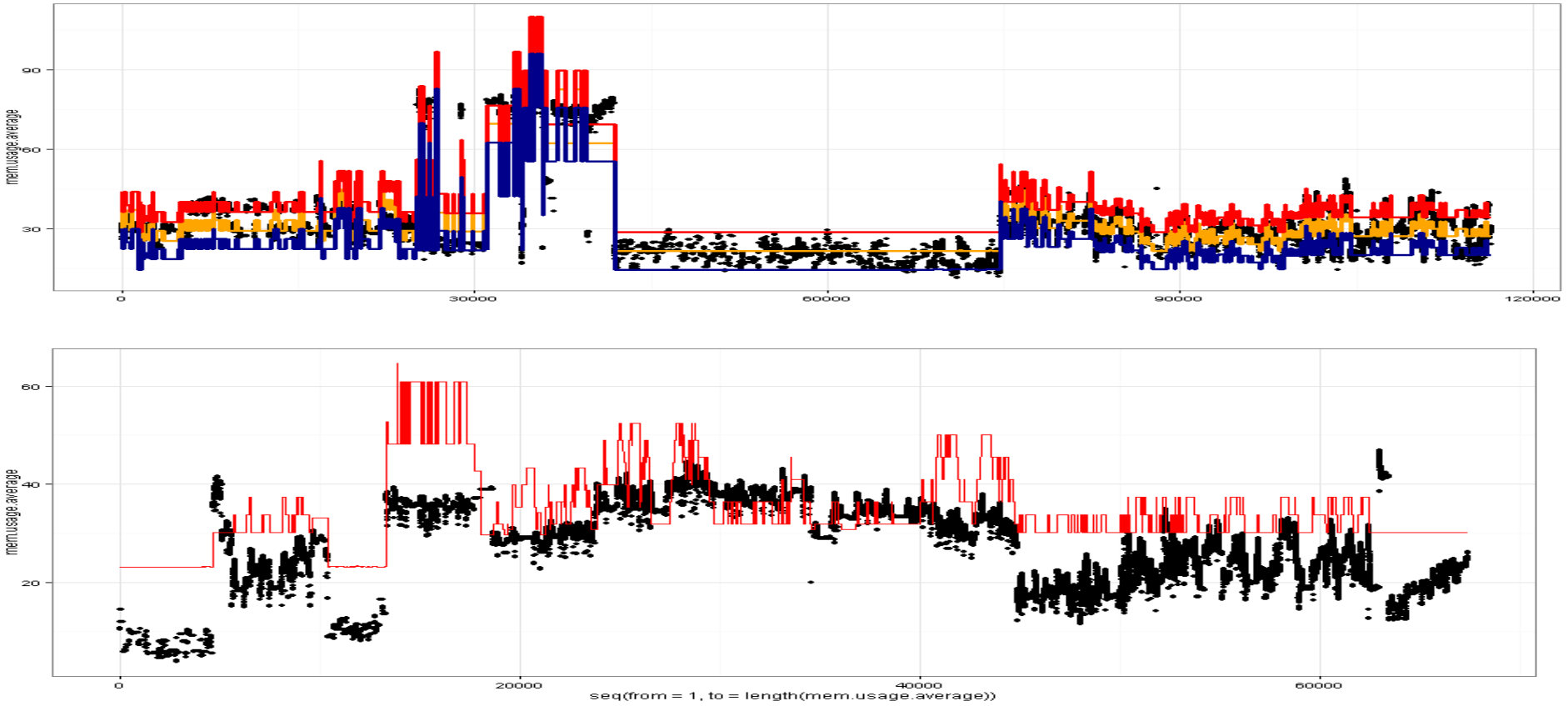
Resource utilization analysis: memory

- Linear model
  - $Mem(VM_1) + Mem(VM_2) + \ldots + Mem(mgmt)$
  - Weighted by the workload



Resource utilization analysis: memory

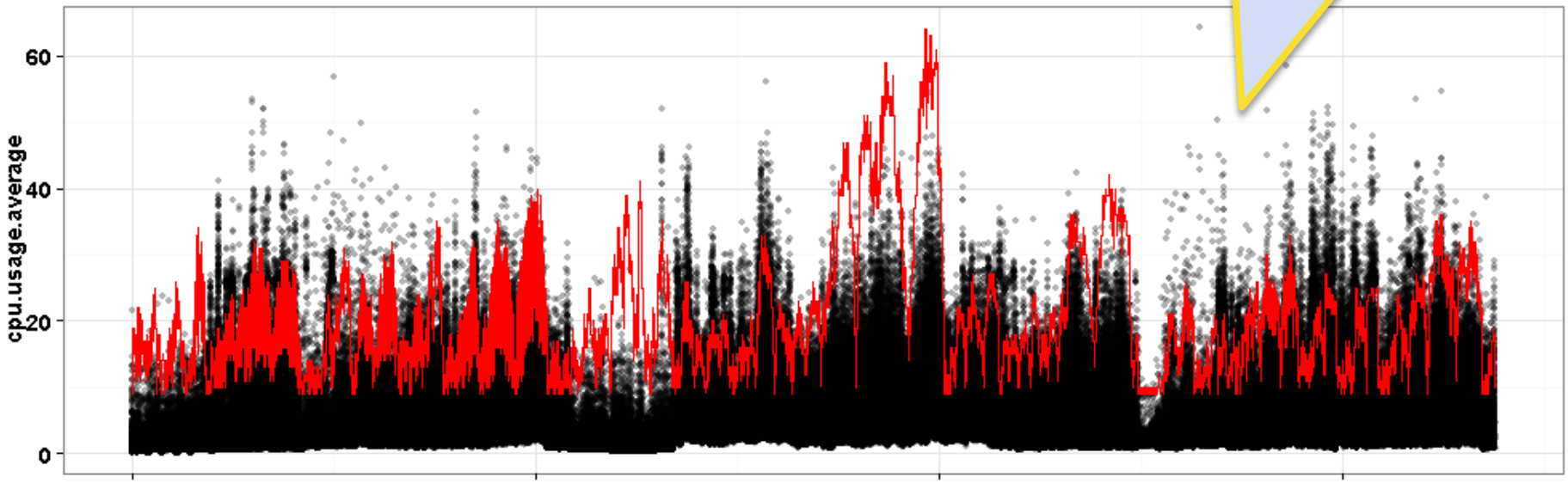
- Linear model
  - $Mem(VM_1) + Mem(VM_2) + \ldots + Mem(mgmt)$
  - Weighted by the workload



Resource utilization analysis: CPU

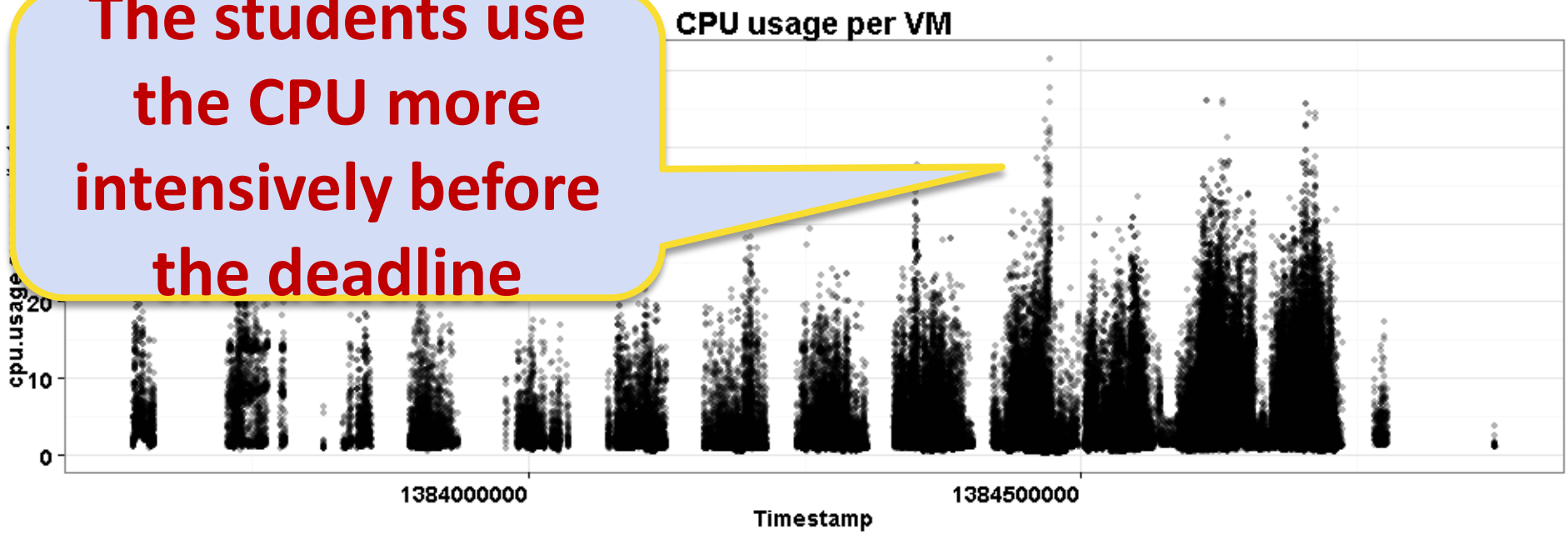
- Linear model
  - $CPU(VM_1) + CPU(VM_2) + \ldots + CPU(mgmt)$
  - Weighted by the workload

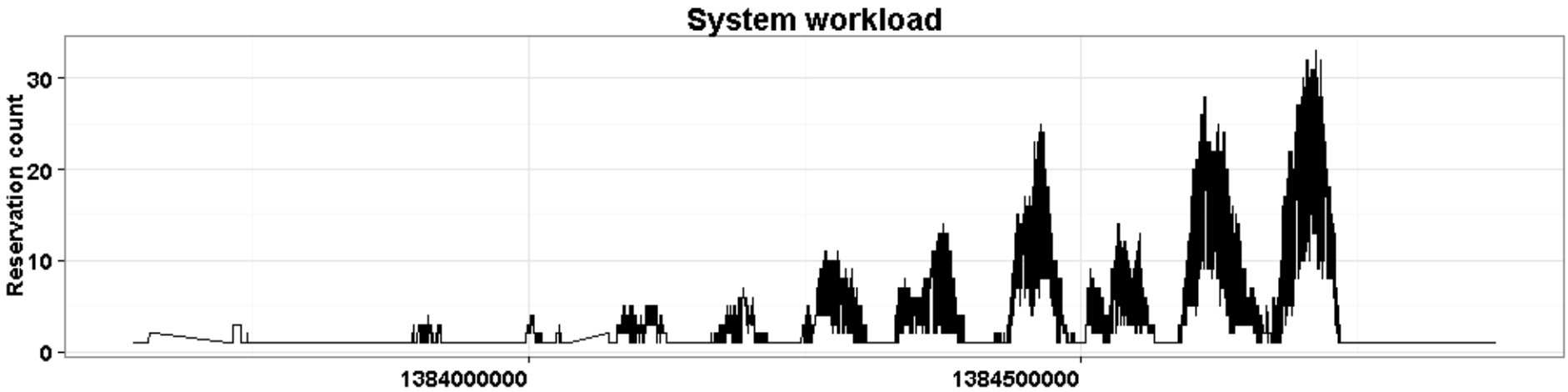
CPU is much more sensitive than memory



Resource utilization analysis: CPU

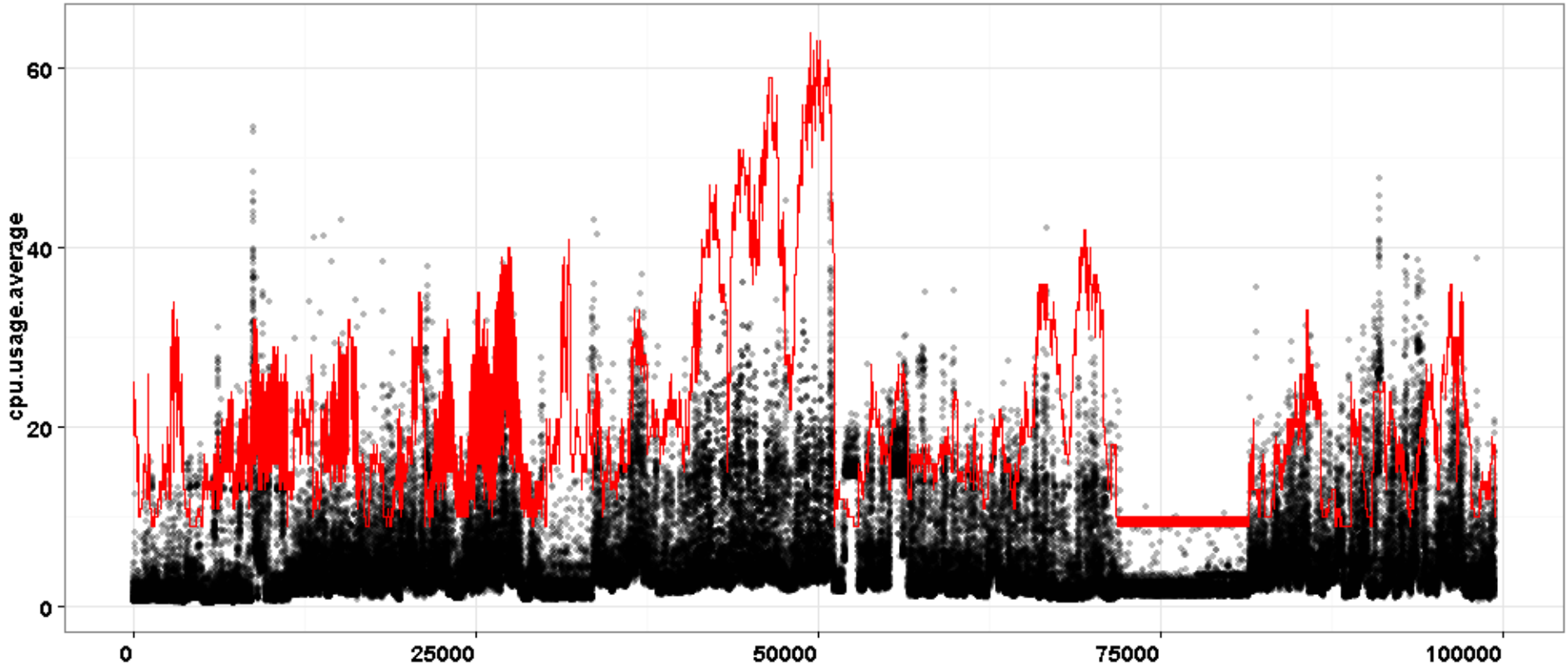
The students use the CPU more intensively before the deadline





Resource utilization analysis: CPU

- Linear model
  - $CPU(VM_1, \boldsymbol{wl}) + CPU(VM_2, \boldsymbol{wl}) + \ldots$
  - Weighted by the workload



Summary

- Data-driven static capacity planning
  - „user behavior” analysis
  - resource fingerprint estimation
- Conclusions:
  - student behavior can be modelled
  - resource allocation were sometimes (too) strict
- Dynamic capacity planning?
  - Long loading time → failed reservations soon
  - When to burst out to a public cloud?