Data Center Specific Thermal and Energy Saving Techniques

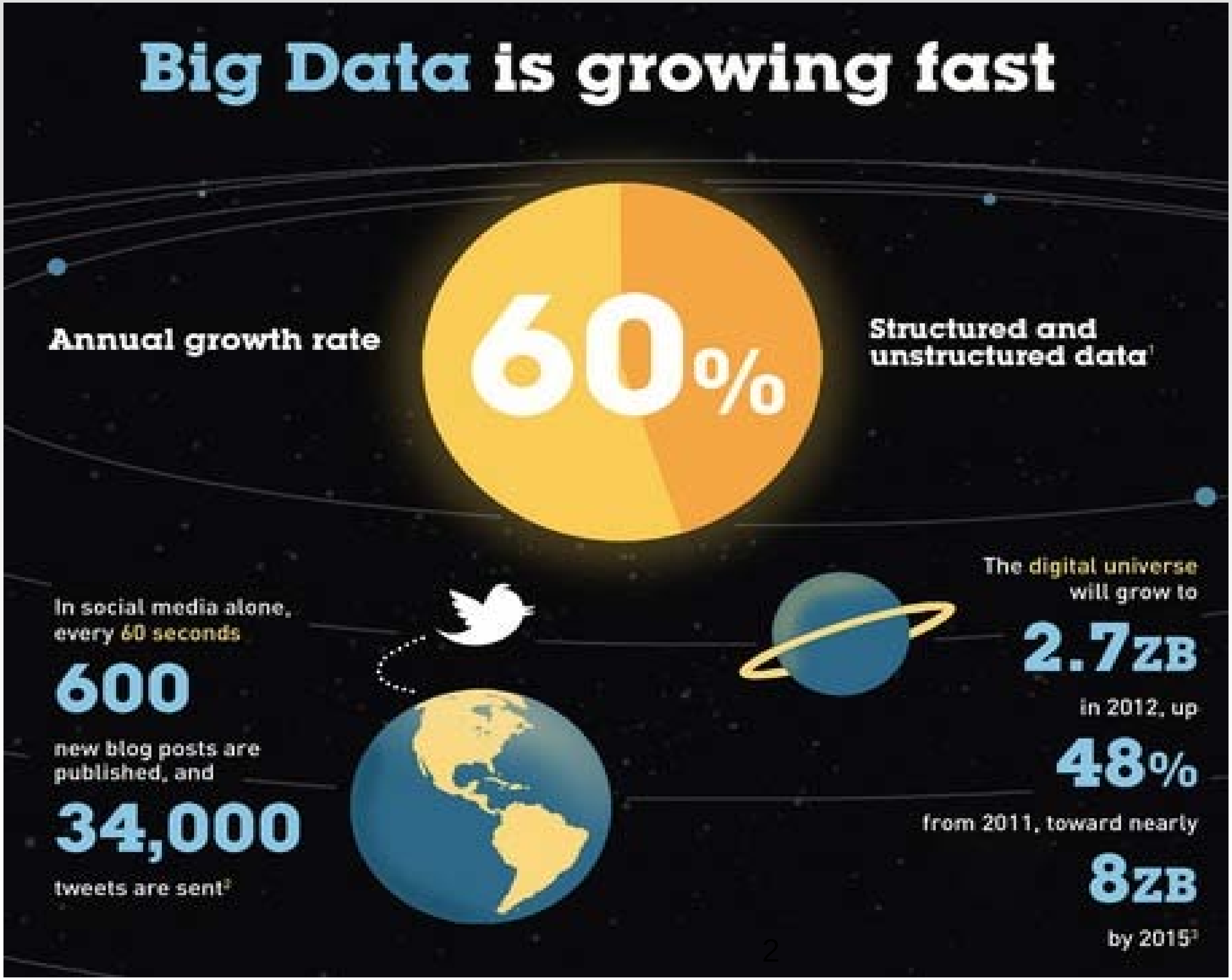
AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

Tausif Muzaffar and Xiao Qin

Department of Computer Science and Software Engineering
Auburn University

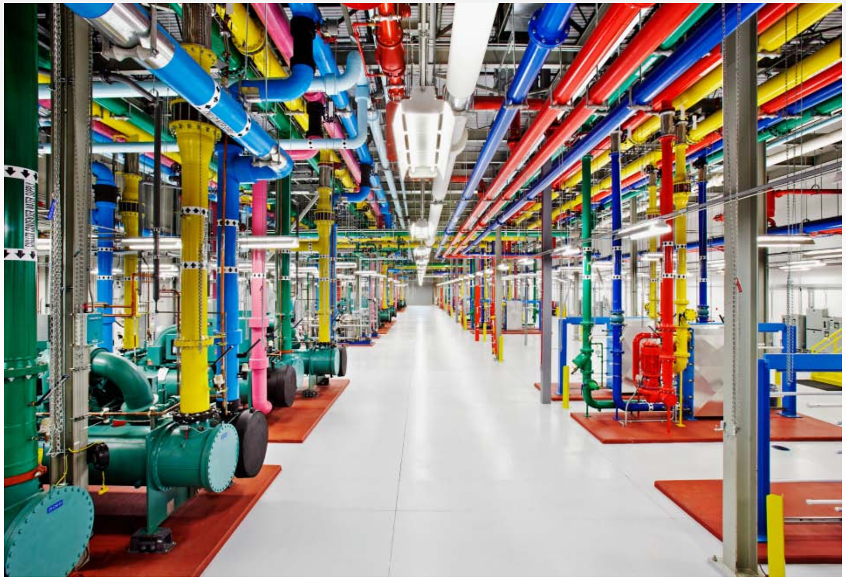
Big Data

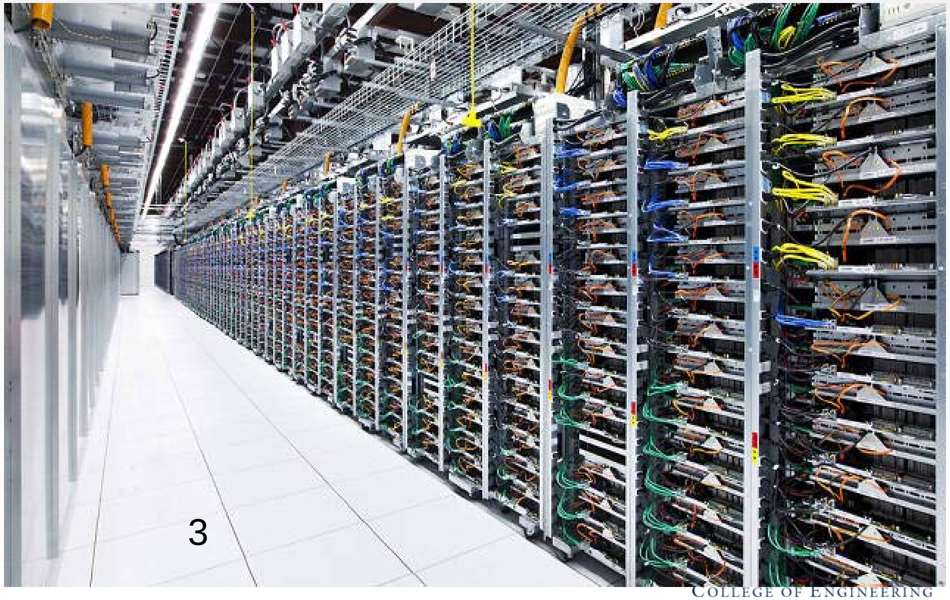


Big Data is growing fast
Annual growth rate
60%
Structured and unstructured data[1]
In social media alone, every 60 seconds
600
new blog posts are published, and
34,000
tweets are sent[2]
The digital universe will grow to
2.7ZB
in 2012, up
48%
from 2011, toward nearly
8ZB
by 2015[3]

Data Centers

- In 2013, there are over 700 million square feet of data centers in united states
- Data centers account for 1.2% of all data power consumed in United States





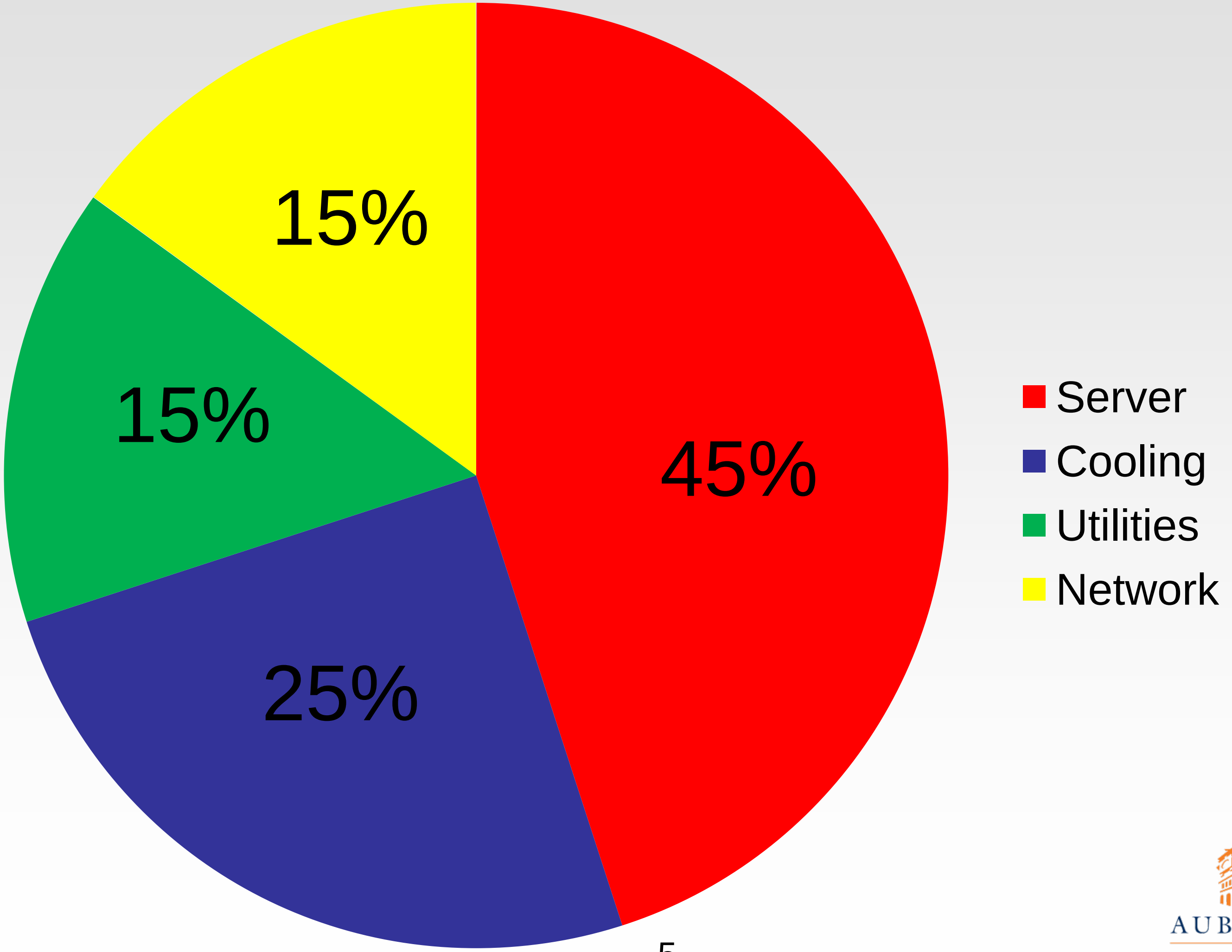
Part 1

THERMAL MODEL

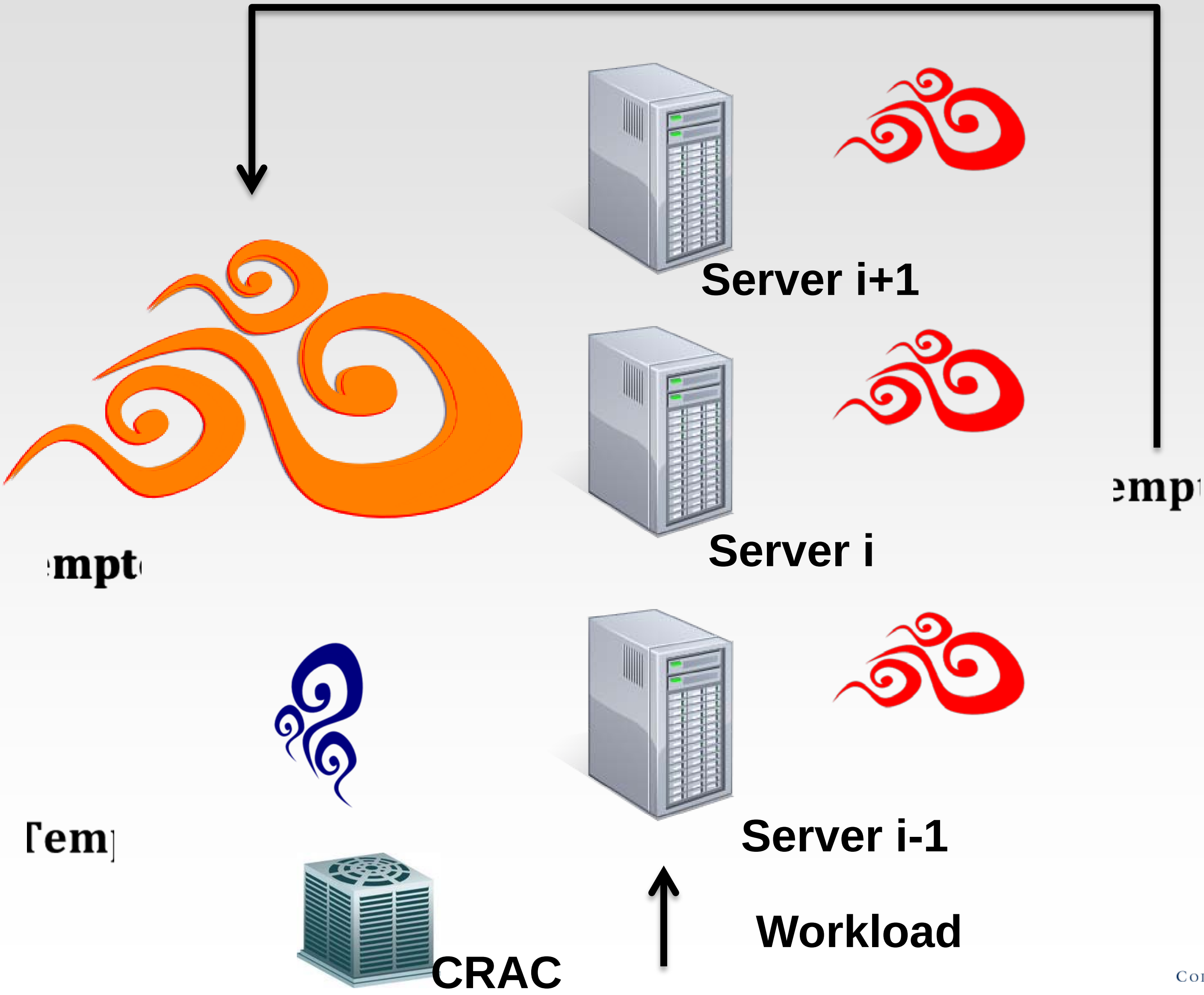
AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

Data Center Power Usage

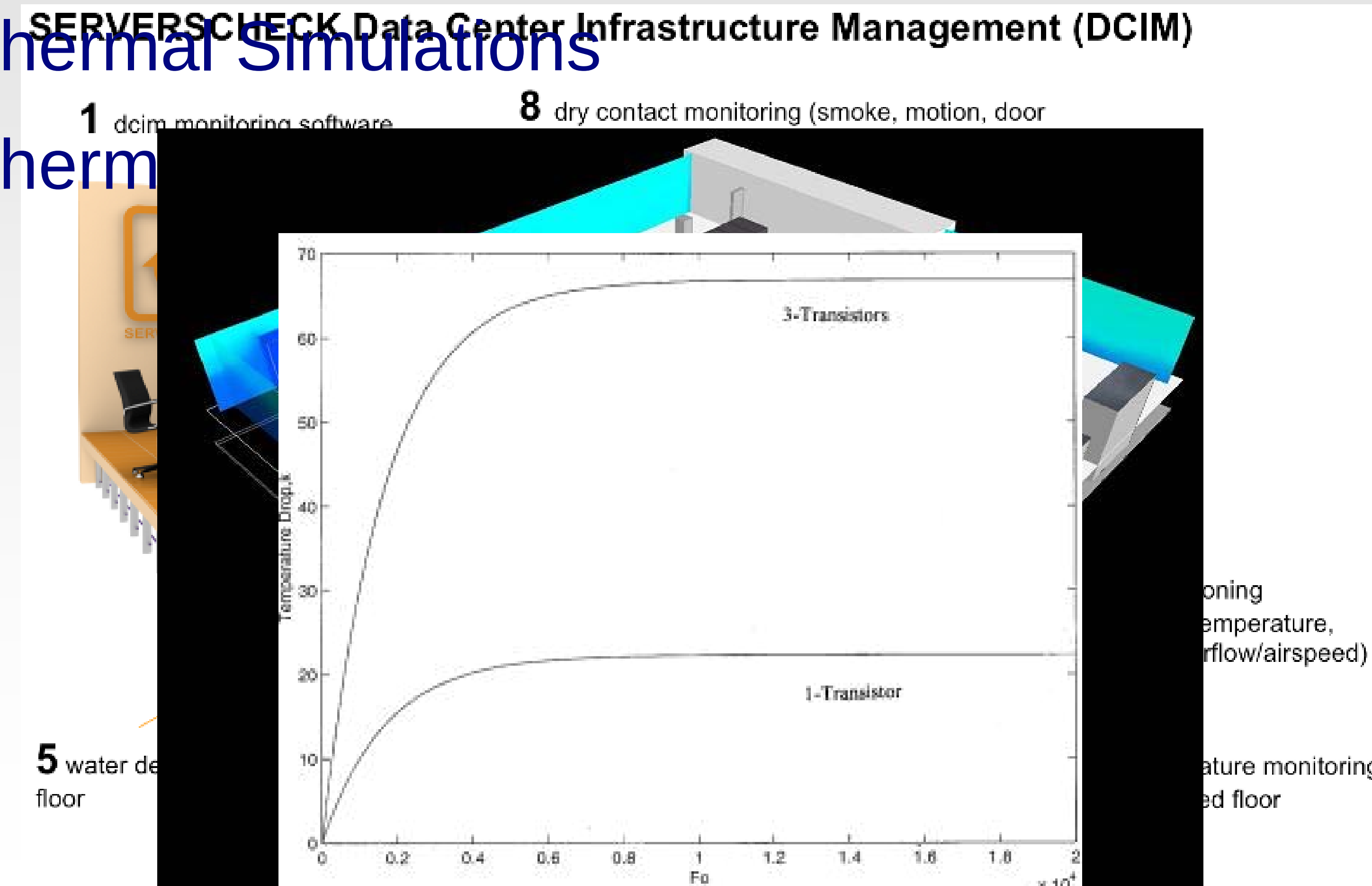


Thermal Recirculation



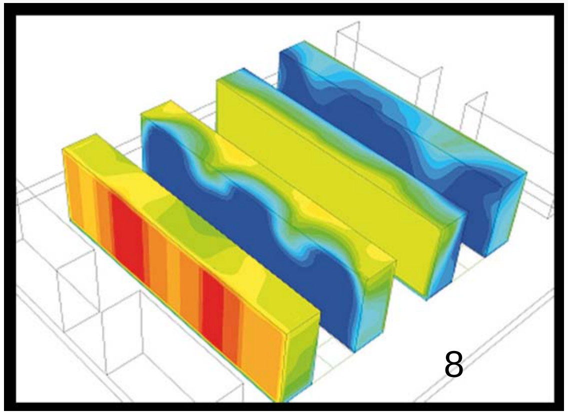
Thermal Recirculation Management

- Sensor Monitoring
- Thermal Simulations
- Therm



Prior Thermal Models

- Some are based on power rather than workload
- Ignore I/O heavy applications
- Requires some sensor support
- Not easily ported to different platforms



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

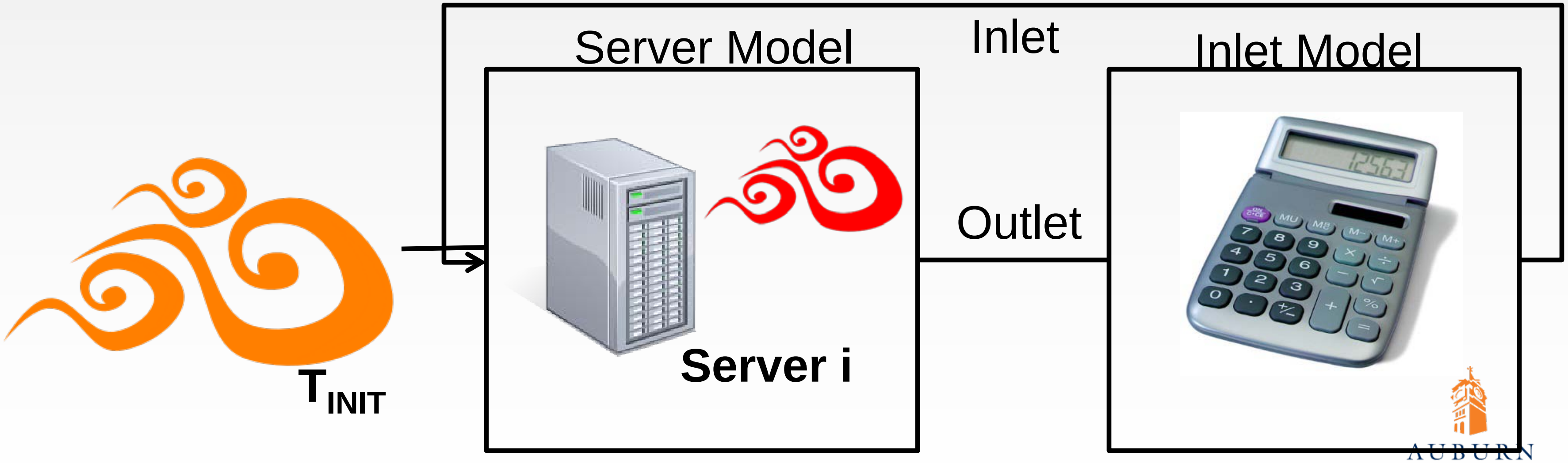
Research Goal

iTad: making a simple and practical way to *estimate the temperature* of a data node based on

- CPU Utilization
- I/O Utilization
- Average Conditions of a Data Center

Our Focus

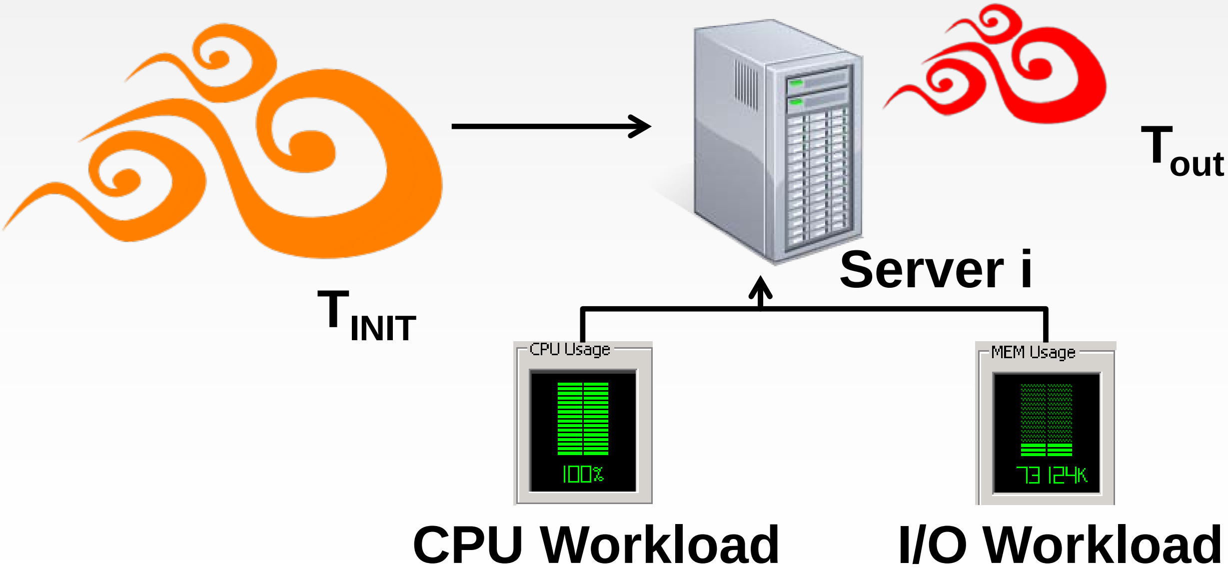
- To focus on each server separately and find the outlet temperature
- To estimate inlet temperature based on that outlet temperature



AUBURN UNIVERSITY
Samuel Ginn College of Engineering

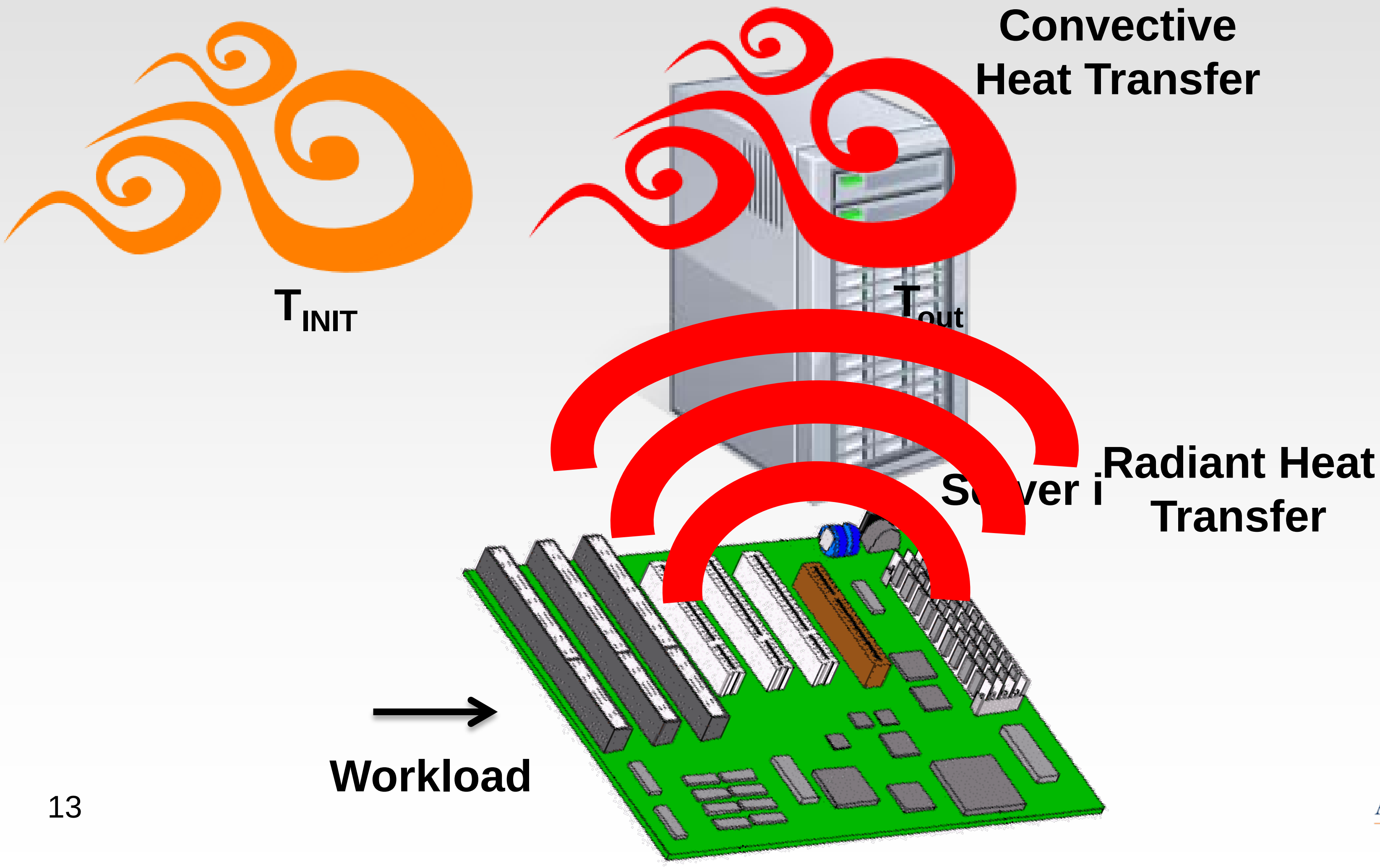
Server Model

- Three factors affect the output temperature of a single node
  - Inlet Temperature
  - CPU Workload
  - I/O Workload



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Server Model Diagram



Server Model Equations

$$Q_i = pfc_p(T_{out_i} - T_{in_i})$$

$$T_{out_i} = \frac{Q_i}{pfc_p} + T_{in_i}$$

(1) Convective Heat Transfer of Server

$$Q_i = h_r A \Delta T_i$$

(2) Radiant Heat Transfer of Server

$$\Delta T_i = \Delta T_{workload_i} + (T_{out_{idle}} - T_{in_{idle}})$$

(3) Change in temperature

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Server Model Equations

$$h_r A \triangle T_i = pfc_p(T_{out_i} - T_{in_i})$$

$$Z = \frac{h_r A}{pfc_p} = \frac{T_{out_i} - T_{in_i}}{\triangle T_i}$$

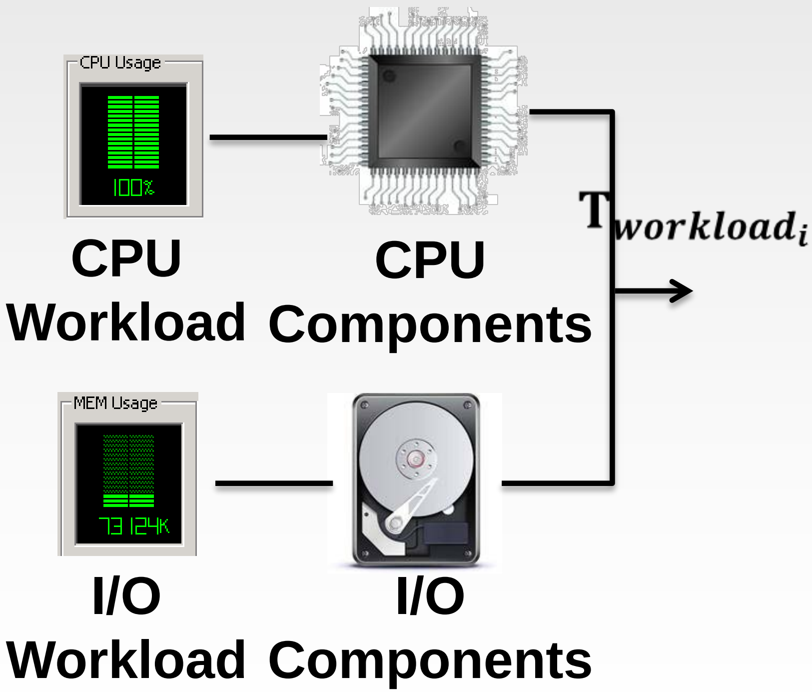
$$T_{out_i} = Z \triangle T_i + T_{in_i}$$

(4) Set Radiant and Convection equal to each other and solve for Tout

AUBURN
UNIVERSITY
SAMUEL GINN
COLLEGE OF ENGINEERING

Workload Model

- To assess how the CPU and I/O effect workload



$$\Delta T_{CPU}^{MAX} = T_{workload_{MAXCPU}} - T_{idle}$$

$$\Delta T_{I/O}^{MAX} = T_{workload_{MAXI/O}} - T_{idle}$$

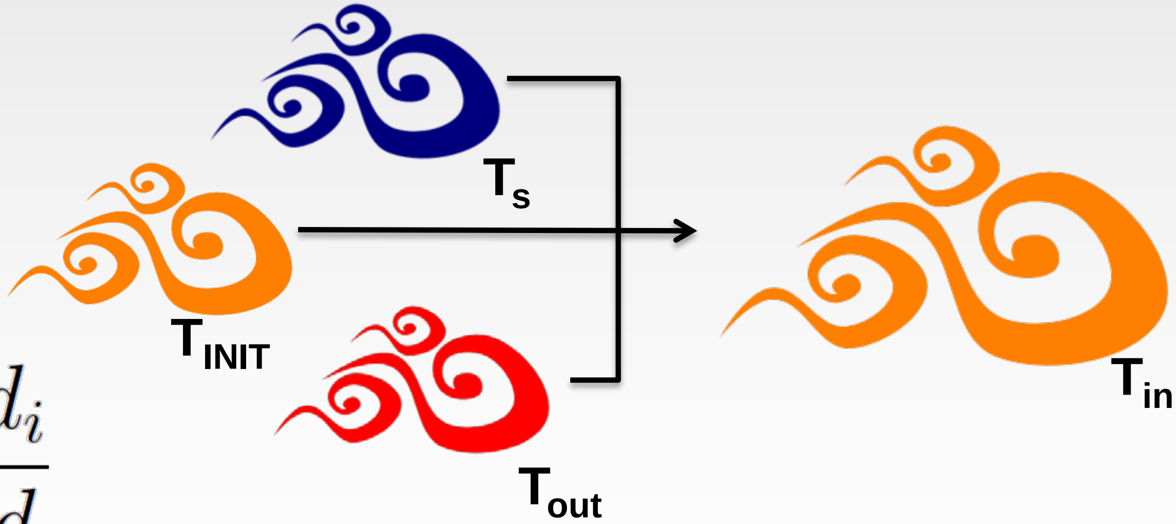
$$\Delta T_{workload_i} = \Delta W_{CPU} \Delta T_{CPU}^{MAX} + \Delta W_{I/O} \Delta T_{I/O}^{MAX}$$

AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

Inlet Model

- After the first run we need to update the inlet temperature to do that we developed this model



$$R = \frac{d_i}{d}$$

$$T_{in} = T_{INIT} - RT_s + kT_{out}$$

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Determining Parameters

- To implement this model we need to get the following constants
  - Maximum I/O and CPU can affect the outlet temperature
  - Z which is a collection of constants

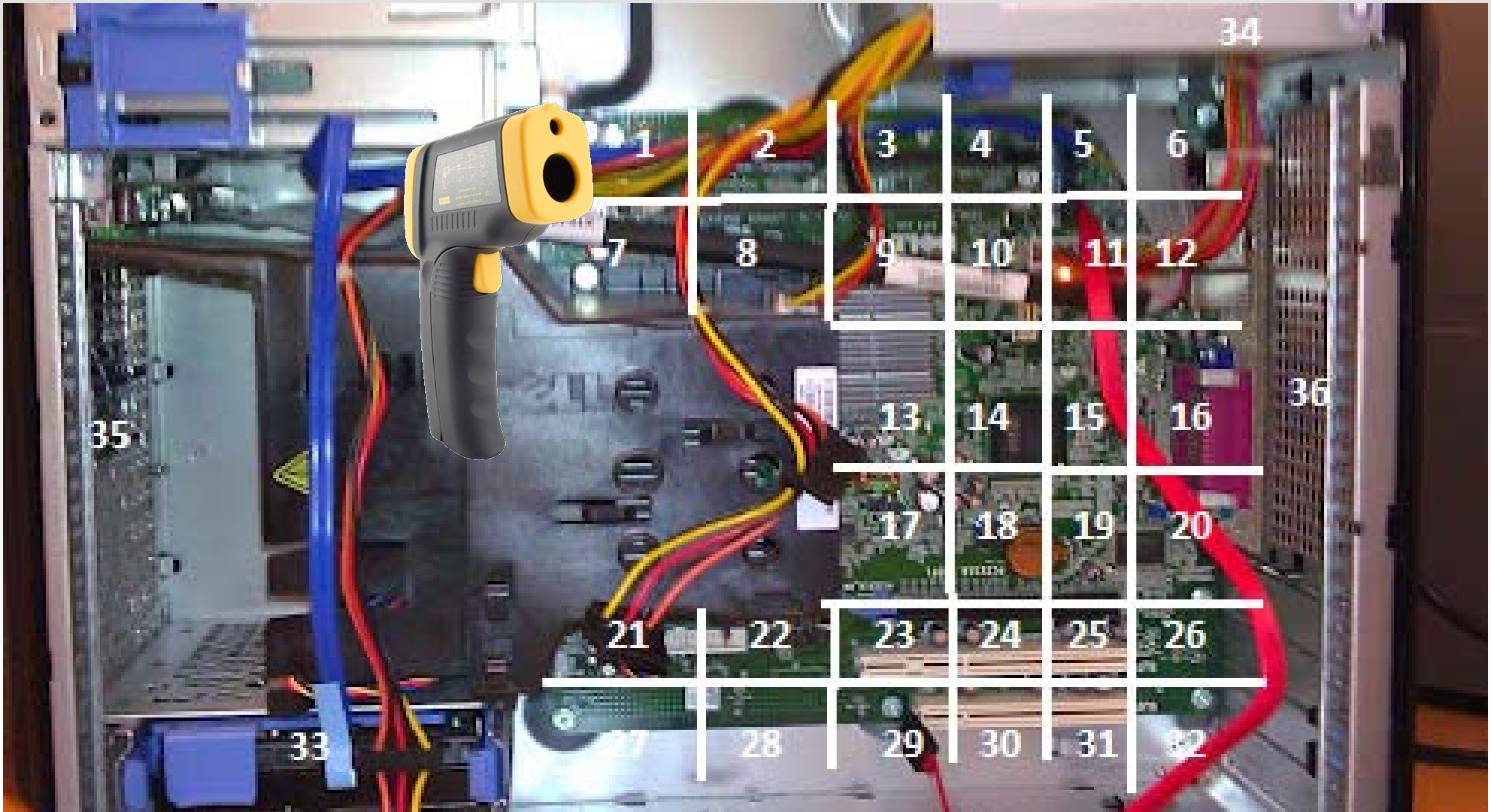
AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Gathering Values

- We thermometers to gather inlet and outlet temperatures
- We used infrared thermometers to get the surface temperature

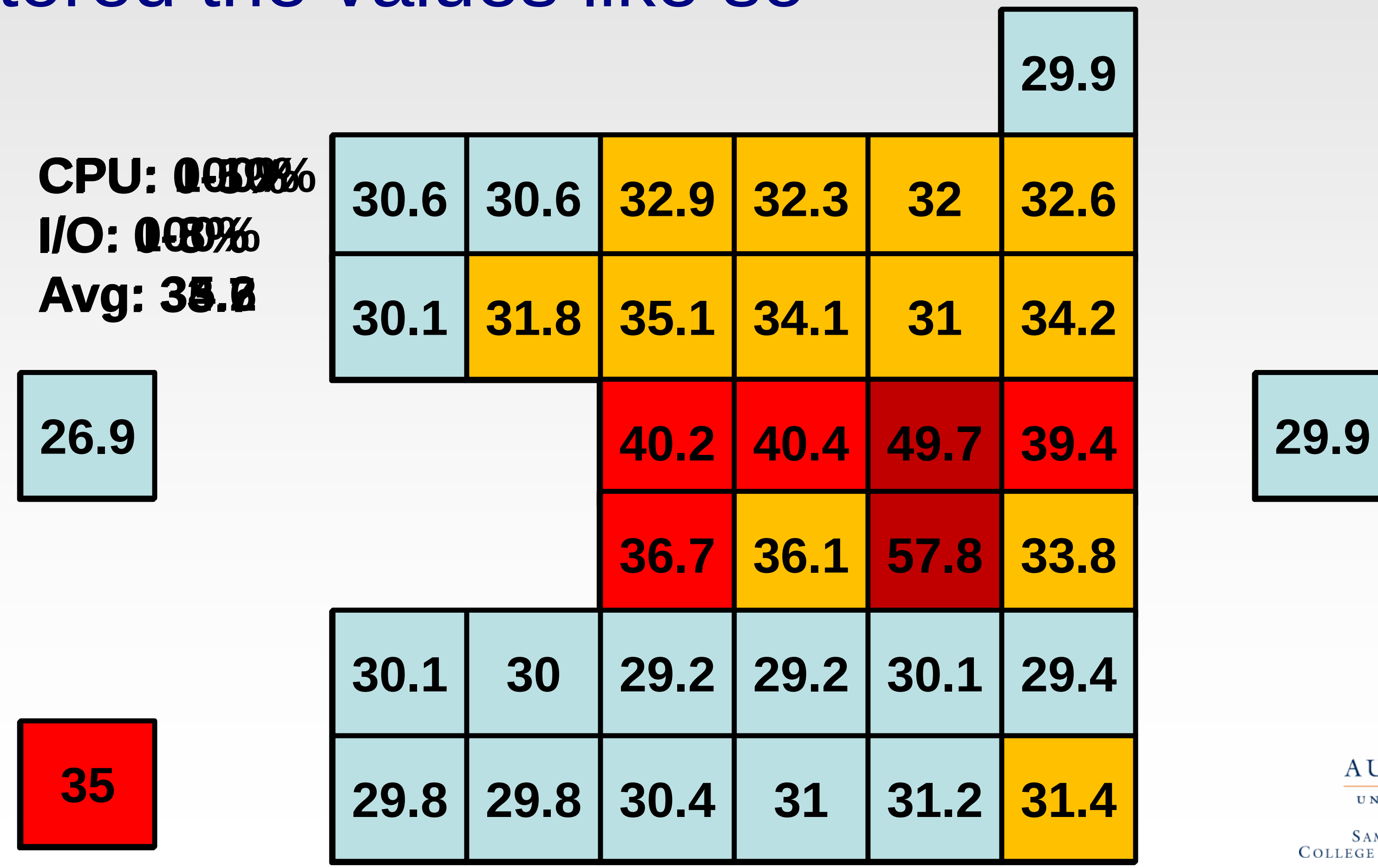
AUBURN
UNIVERSITY
SAMUEL GINN
COLLEGE OF ENGINEERING

Test Machines



Data Capture

- We gathered surface temperature and stored the values like so

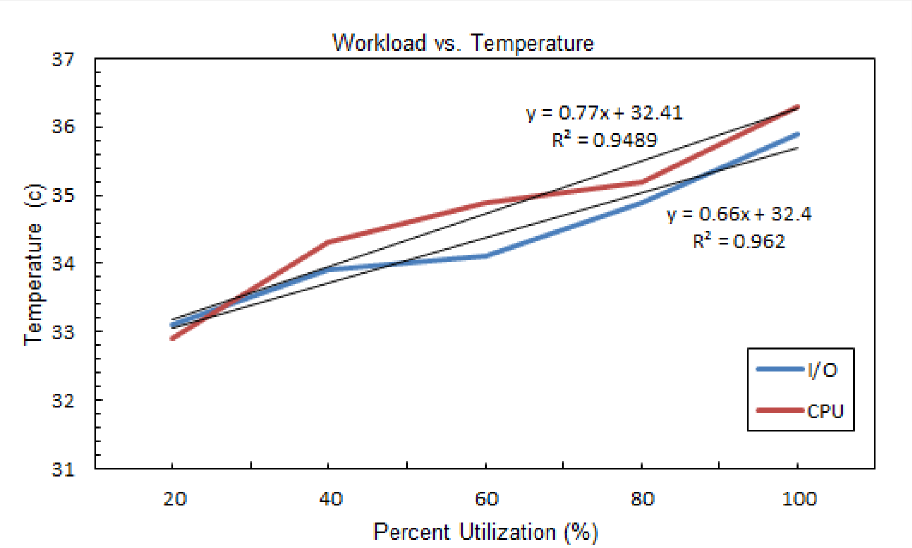


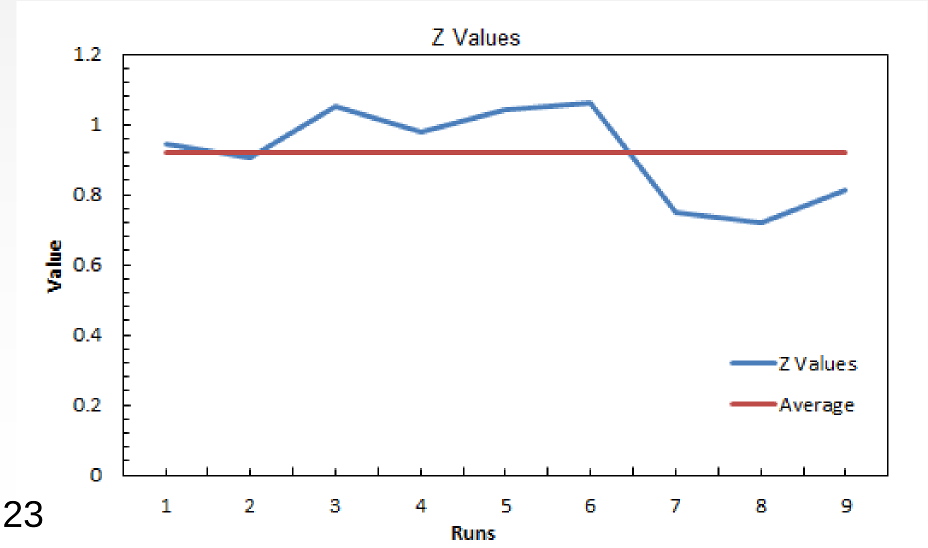
AUBURN UNIVERSITY
Samuel Ginn College of Engineering

Determining Constants

- We observed the rate in changed with CPU and I/O
- We used the values to calculate Z

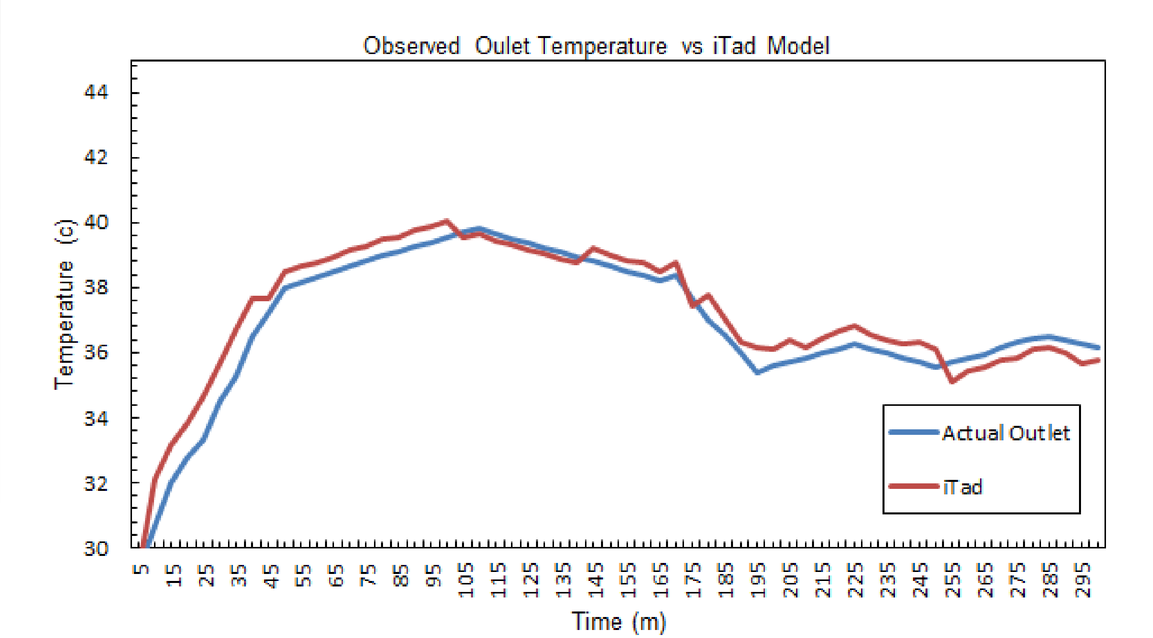
$$T_{out_i} = Z \Delta T_i + T_{in_i}$$





Verification

- After getting the constants we ran a live test where we had a computer run tasks and we measured actual outlet temperatures vs. model outlet temperature



AUBURN UNIVERSITY
Samuel Ginn College of Engineering

Implementations

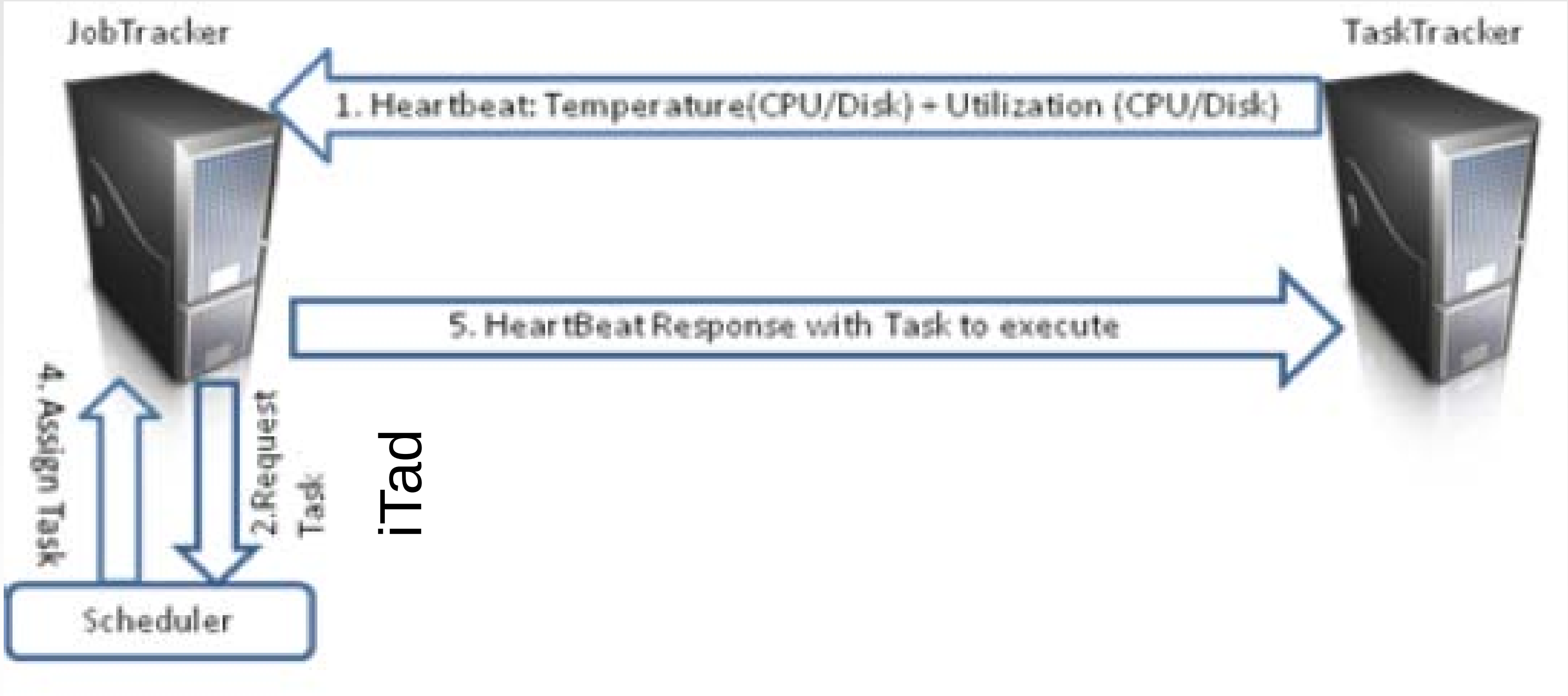
- MPI using iTad to decisions

```
if(iTad() < 29)
{
    if(myid != 0)
    {
        MPI_Recv(&number, 1, MPI_INT, myid-1, 0, MPI_COMM_WORLD,MPI_STATUS_IGNORE);
    } else {
        MPI_Recv(&number, 1, MPI_INT, world_size, 0, MPI_COMM_WORLD,MPI_STATUS_IGNORE);
    }
} else {
    while(iTad() > 29){
        sleep(10);
    }
    if(myid != 0)
    {
        MPI_Recv(&number, 1, MPI_INT, myid-1, 0, MPI_COMM_WORLD,MPI_STATUS_IGNORE);
    } else {
        MPI_Recv(&number, 1, MPI_INT, world_size, 0, MPI_COMM_WORLD,MPI_STATUS_IGNORE);
    }
}
```


AUBURN
UNIVERSITY
SAMUEL GINN
COLLEGE OF ENGINEERING

Implementations

- We added iTad to Hadoop Heartbeat



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Part 2

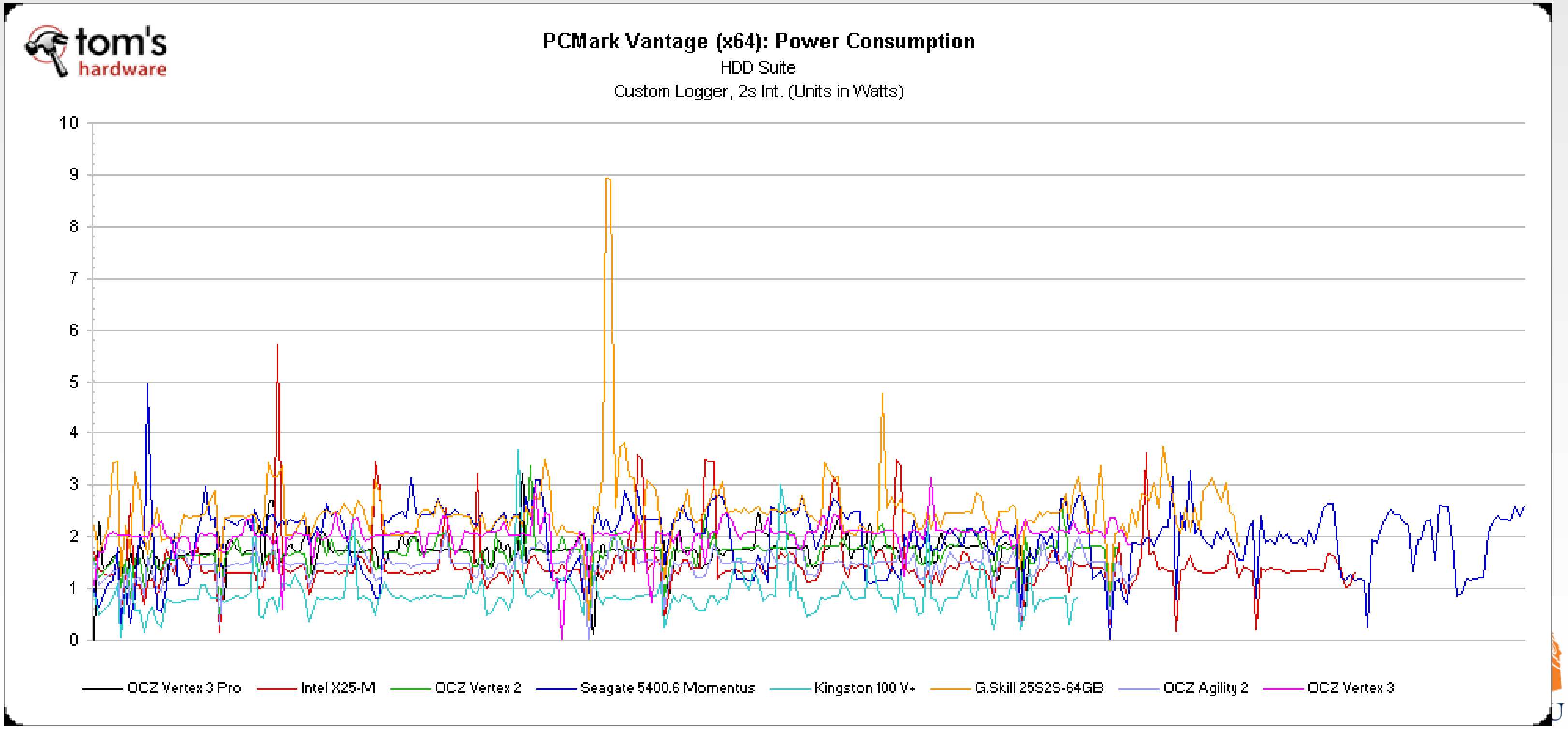
HADOOP DISK ENERGY EFFICIENCY

AUBURN UNIVERSITY

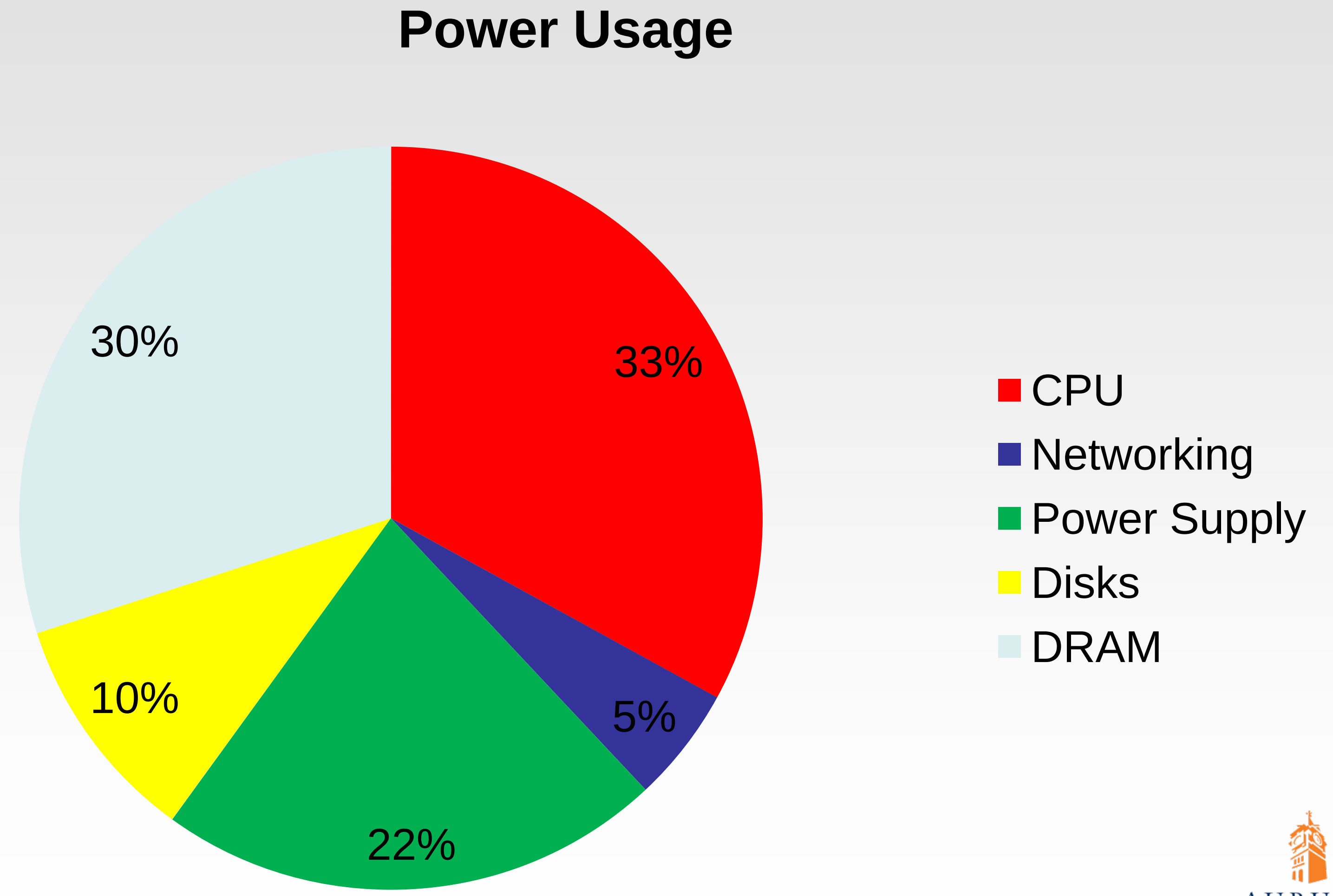
SAMUEL GINN COLLEGE OF ENGINEERING

Disk Energy

- Disk drives varies in energy
- Disks can be a significant part of a server



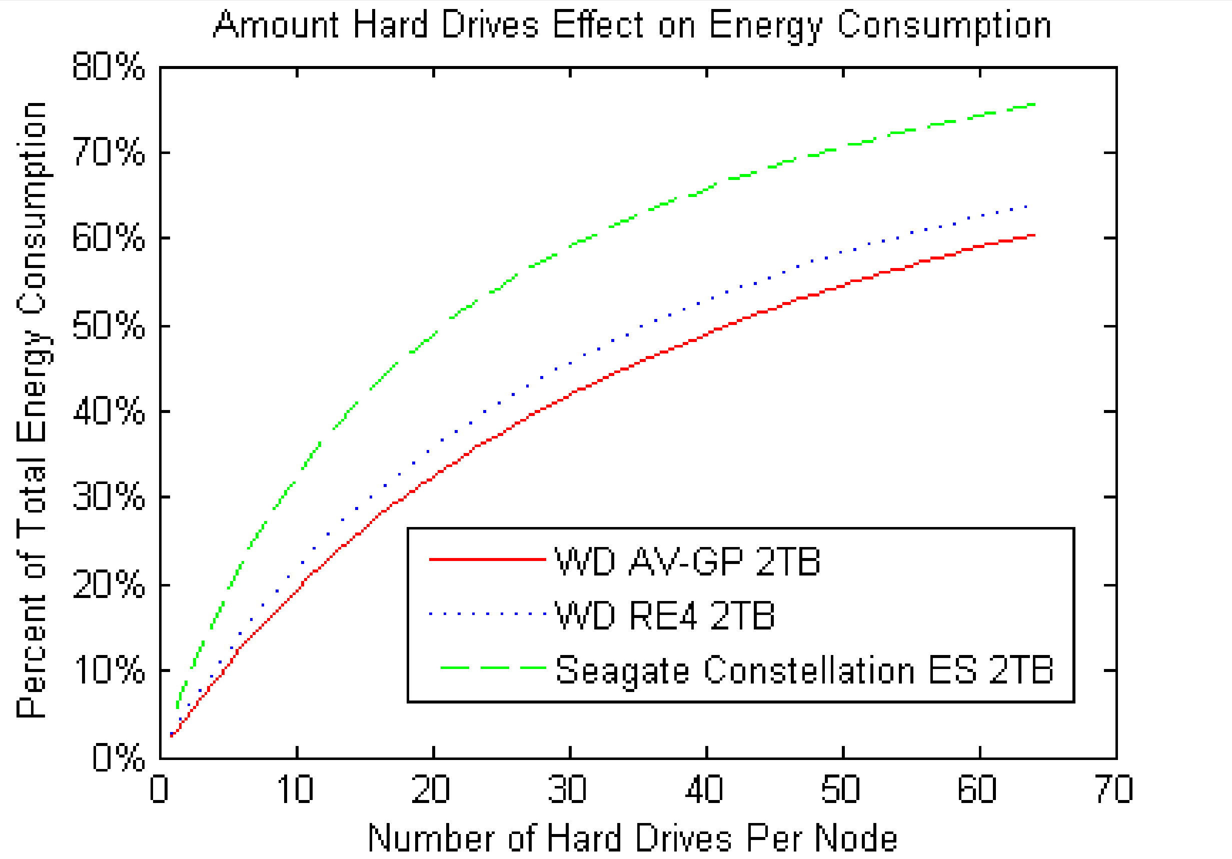
Single-Disk Server Power Usage



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Scaling Server Disk

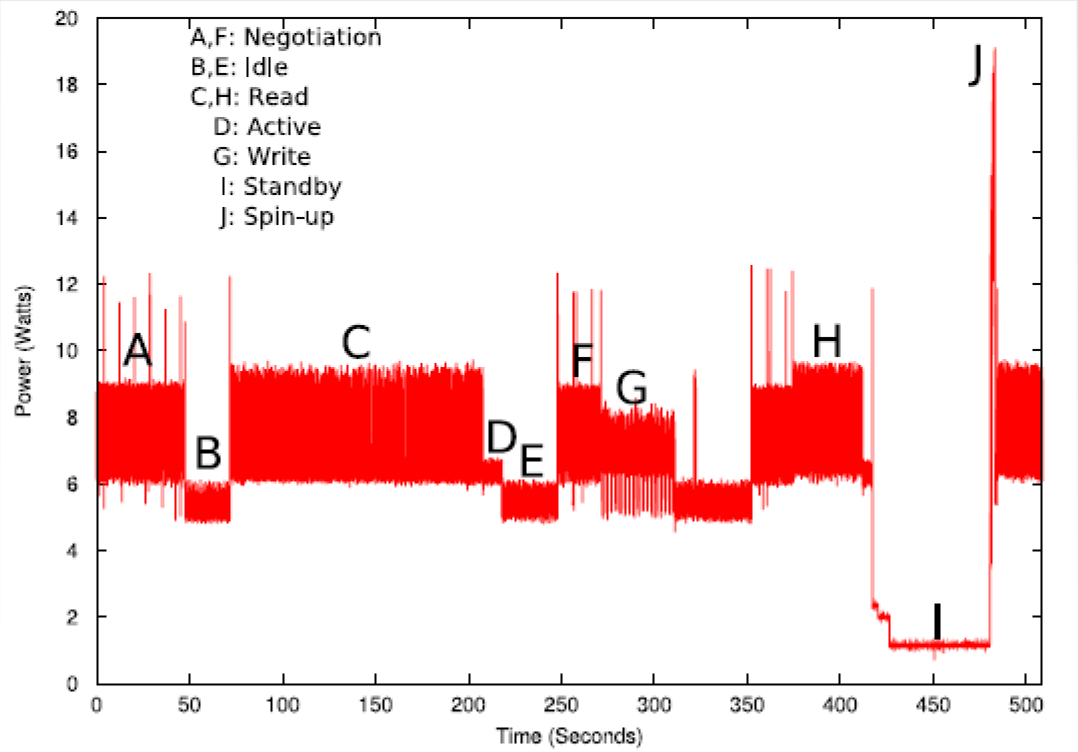
- With every added disk, hard drive energy plays a bigger role



AUBURN UNIVERSITY
Samuel Ginn College of Engineering

Disk Dynamic Power

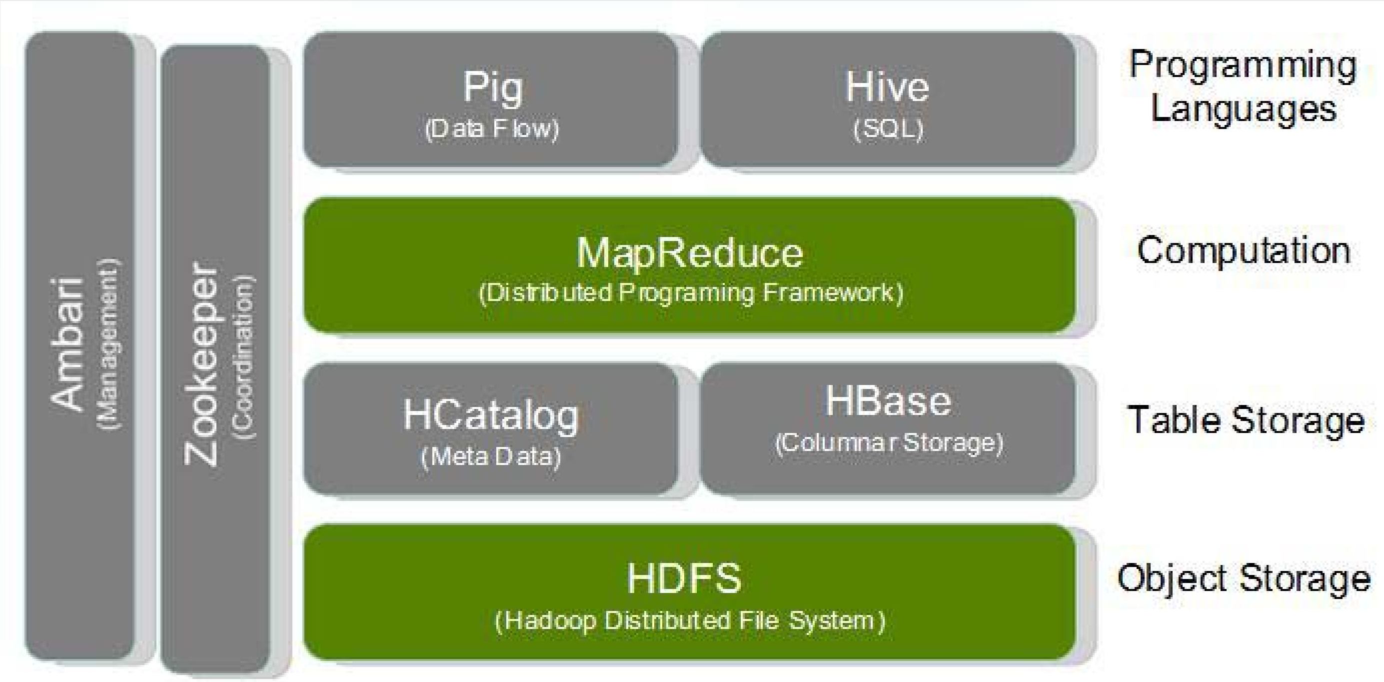
- Disks tend to have different consumption modes
  - Active
  - Idle
  - Standby



AUBURN UNIVERSITY
Samuel Ginn College of Engineering

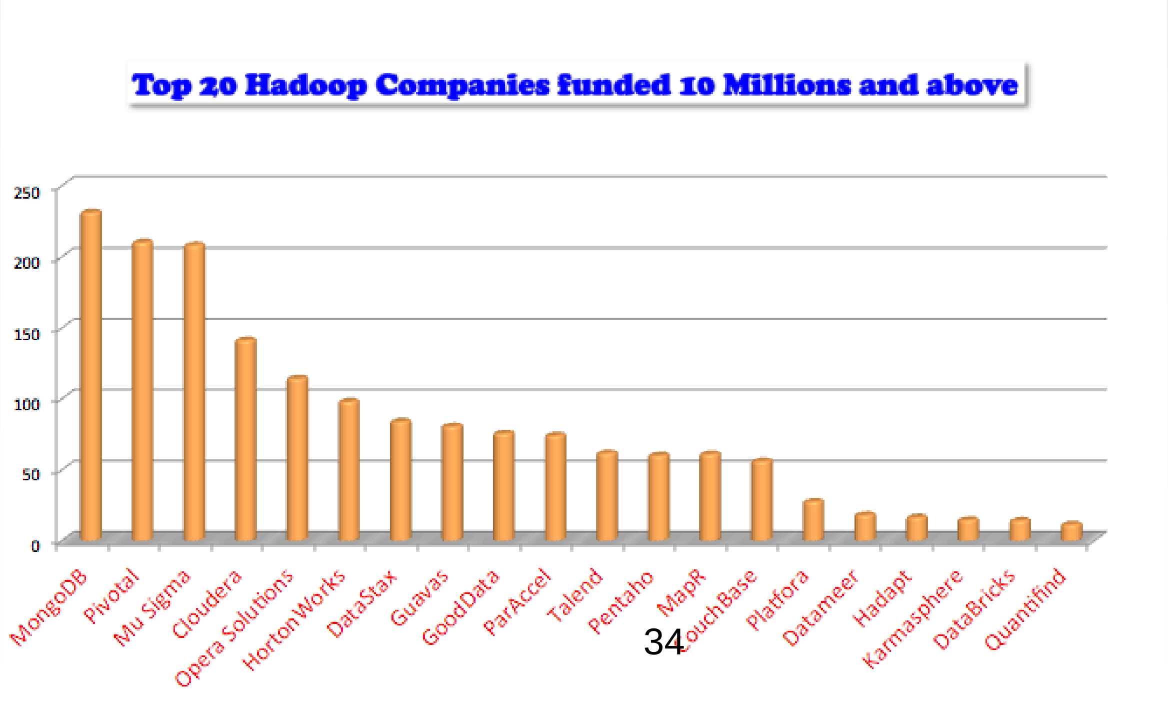
Hadoop Overview

- Parallel Processing
  - Map Reduce
- Distributed Data



Hadoop Benefits

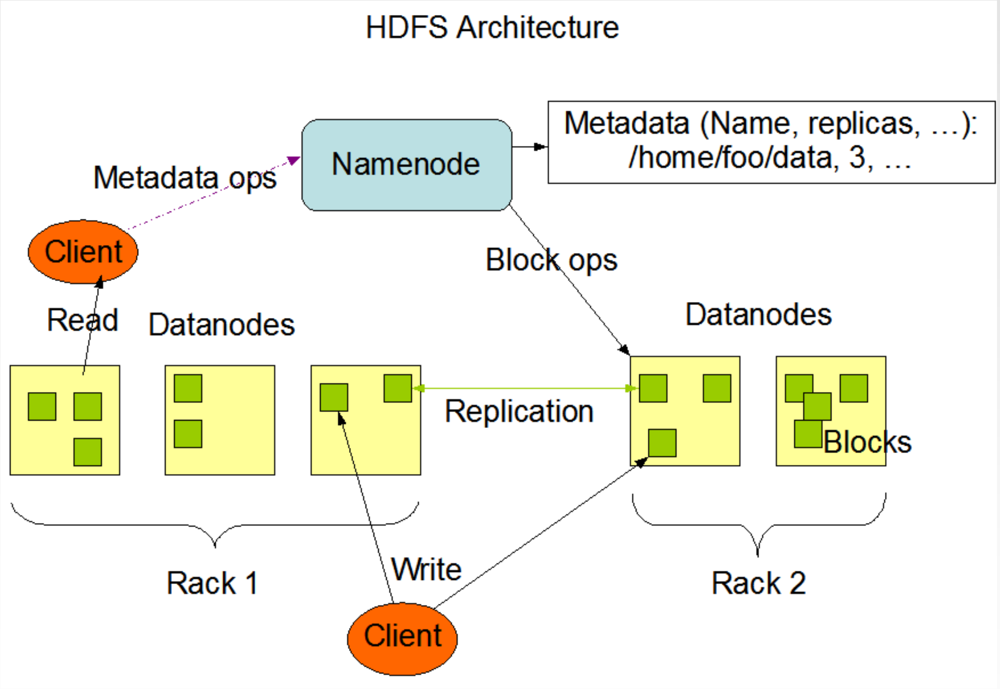
- Industry Standard
- Large Research Community
- I/O Heavy



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Hadoop Architecture

- Hadoop creates multiple replicas
- Metadata is managed on name node
- Nodes can have multiple disks



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Research Goal

NAP – E(N)ergy (A)ware Disks for Hadoo(P)

- Built for high energy efficiency
- Designed for Hadoop clusters

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

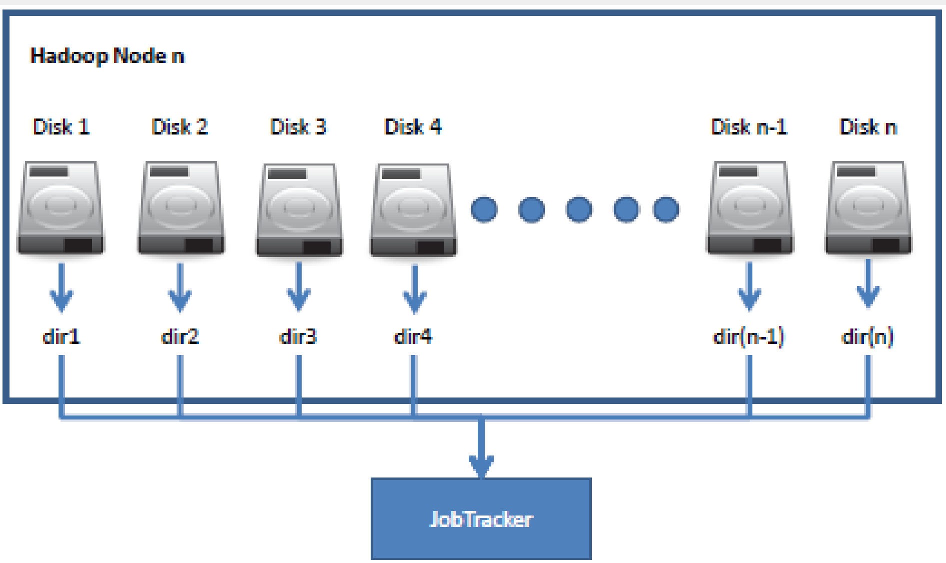
Setup

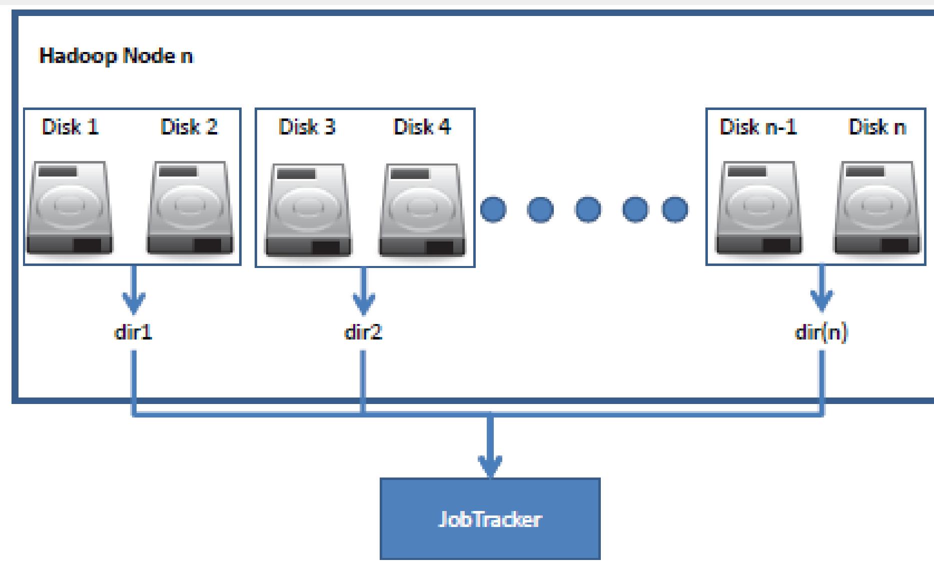
- 3-node cluster
- Each node identical
  - 4 disks
  - 4gb RAM
- Cloudera Hadoop
- Power meter

AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Optimizations

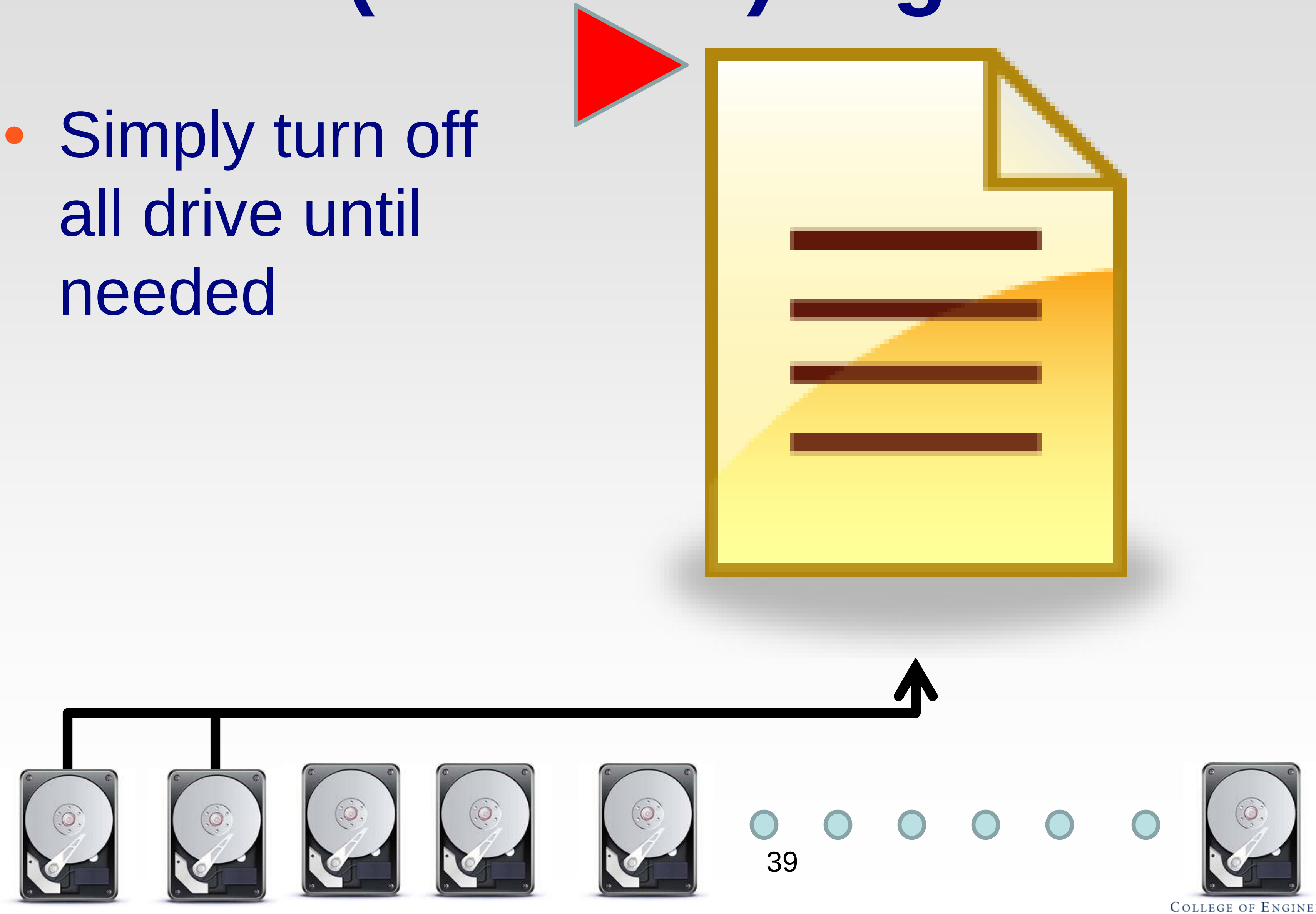
- We group disks together
  - I/O Limits
  - More time for disks to sleep





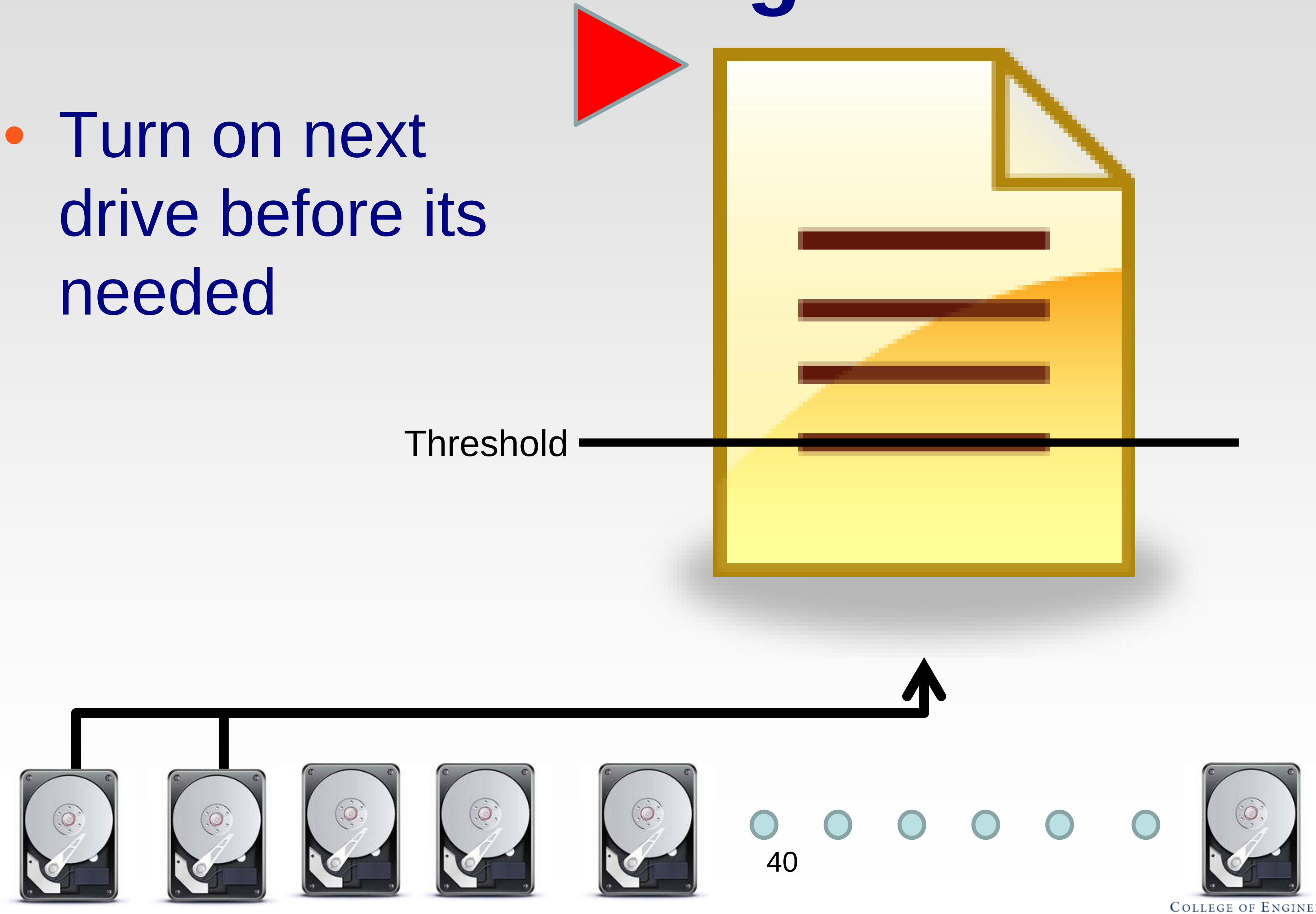
Naïve (Reactive) Algorithm

- Simply turn off all drive until needed



Proactive Algorithm

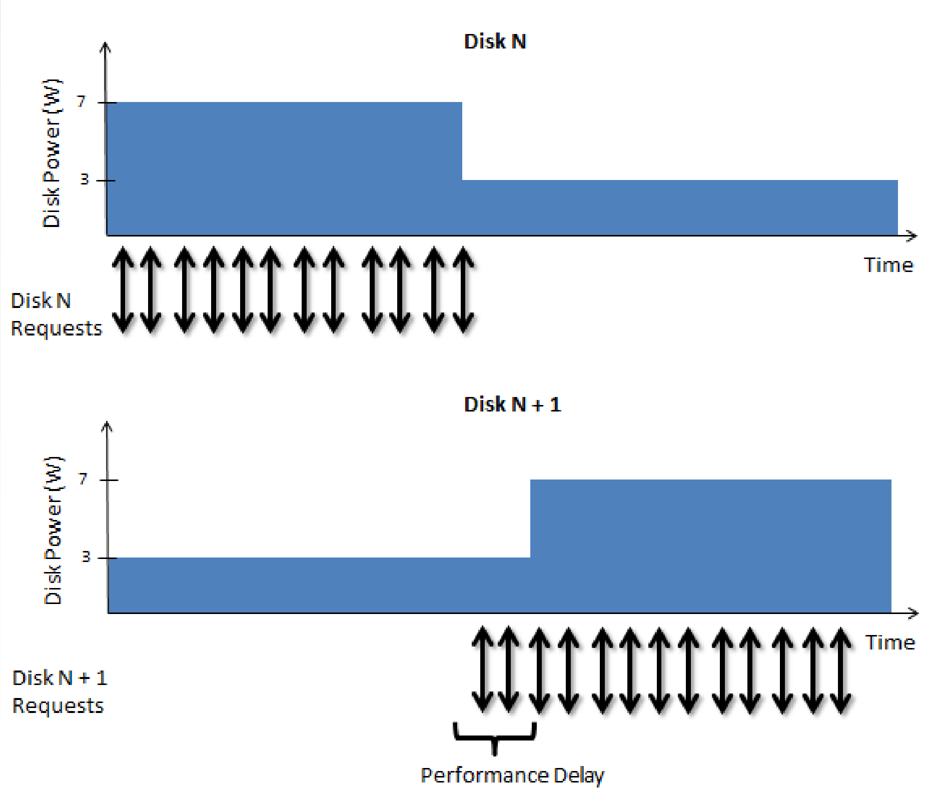
- Turn on next drive before its needed



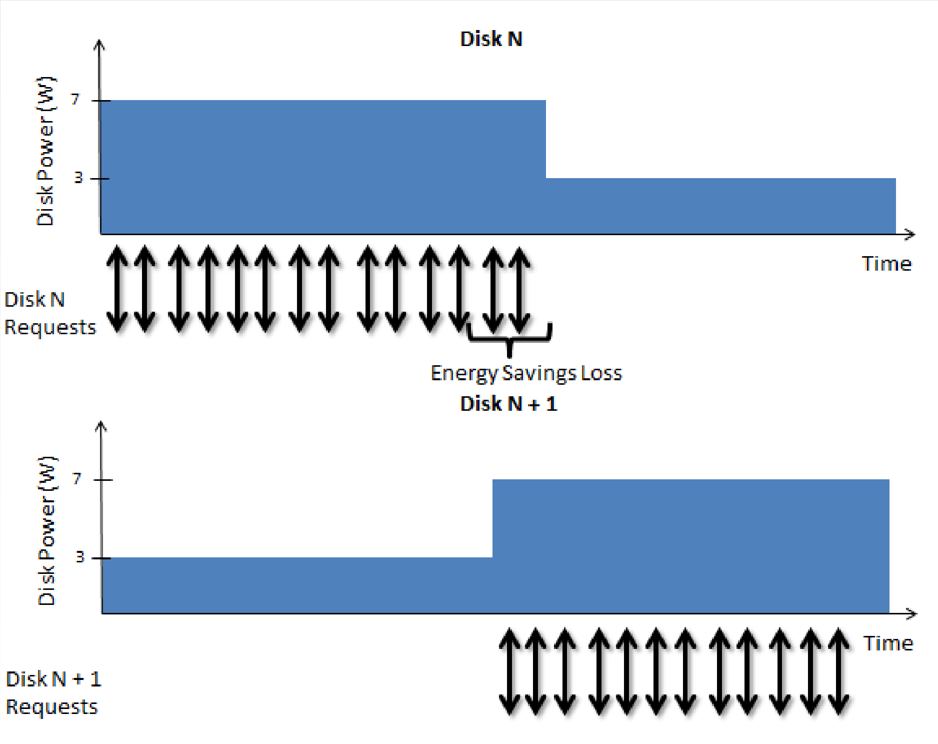
College of Engineering

Comparing the Algorithms

Reactive



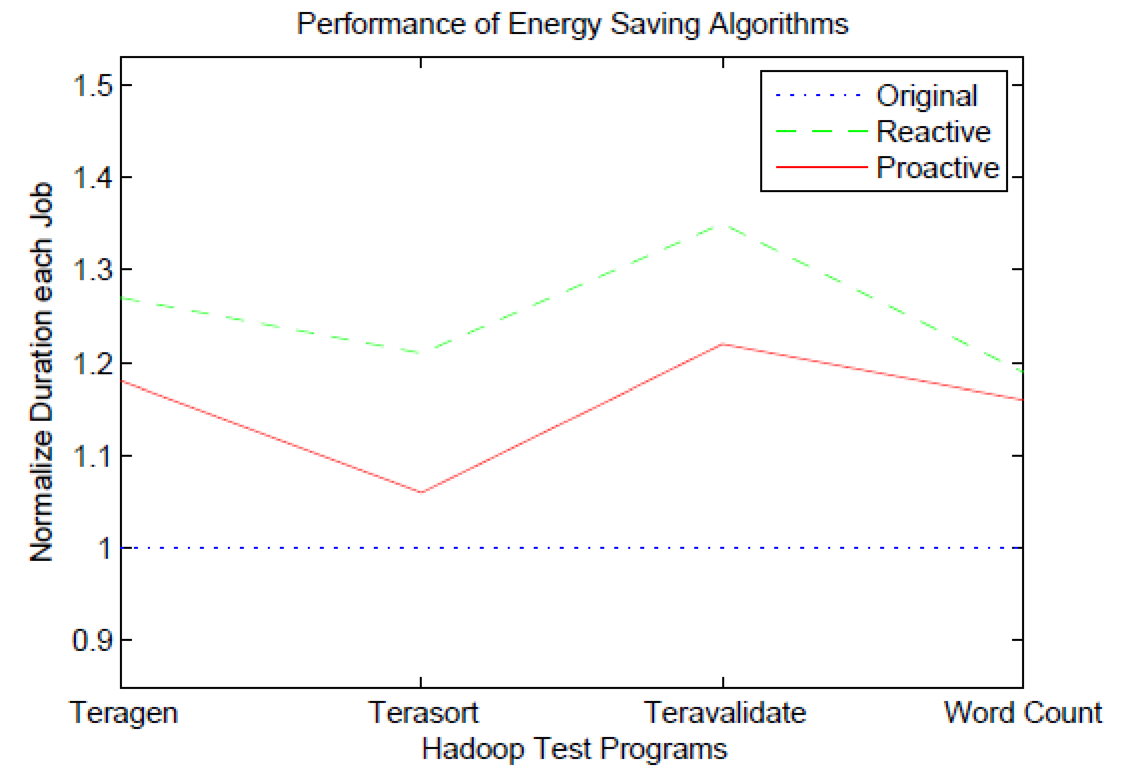
Predictive



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Speed

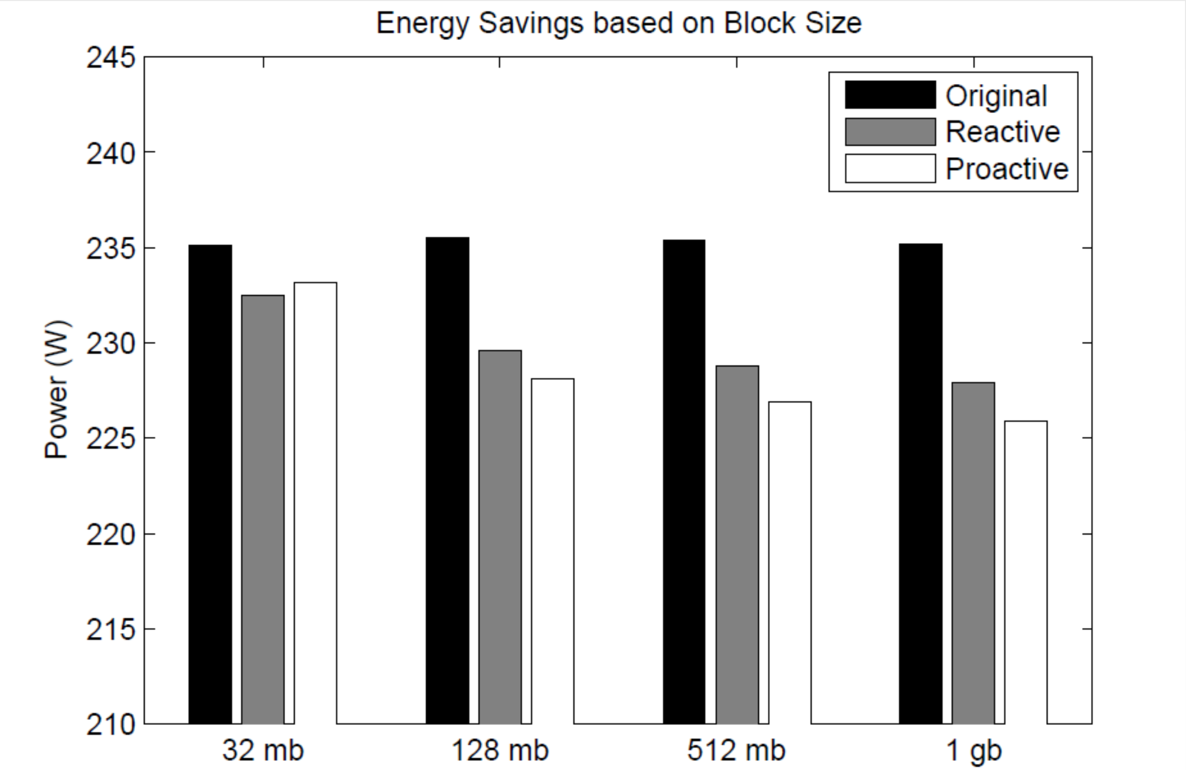
- Reactive does worse than proactive
- Time increase low



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Block Size

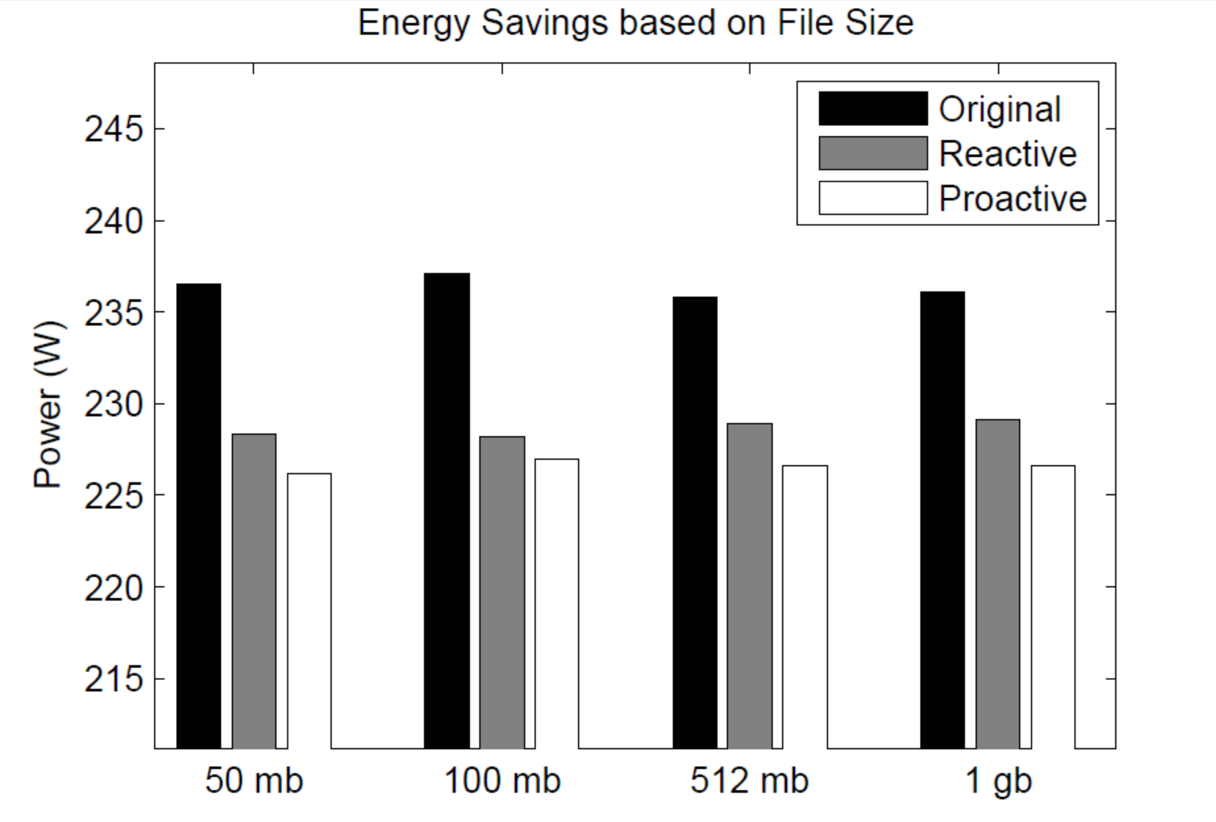
- Effects how HDFS stores files
- Effects how fast it processes



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

File Size

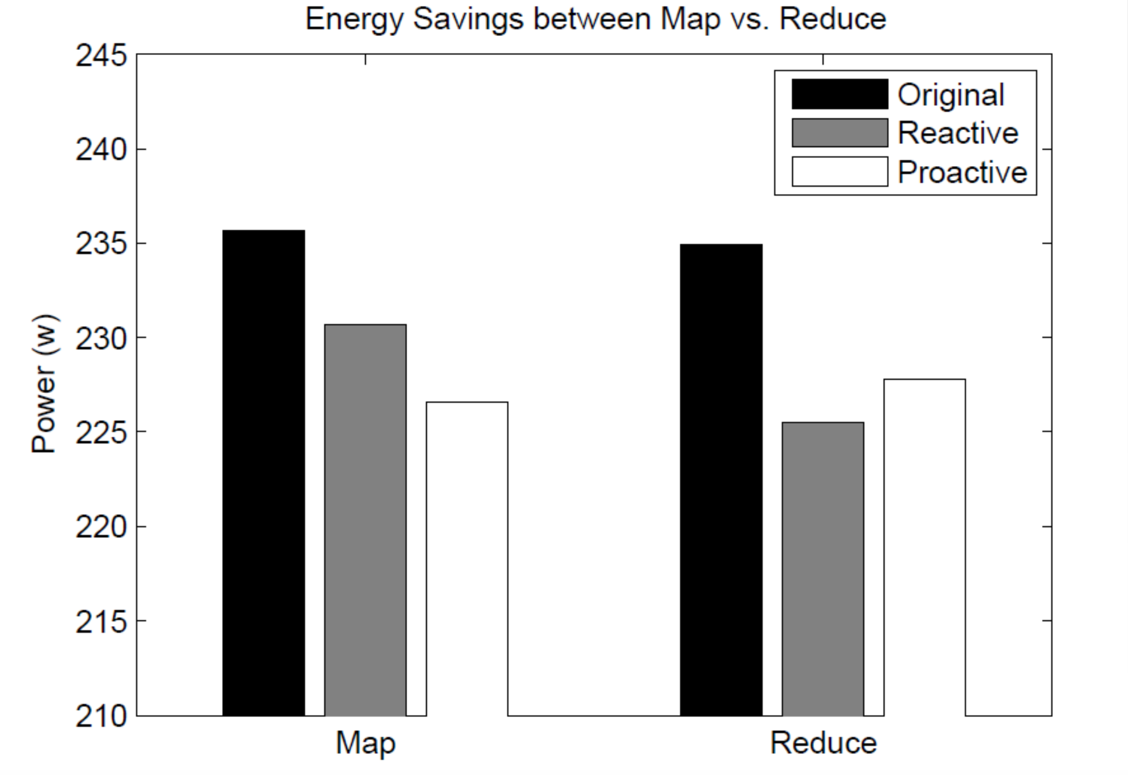
- Effects how blocks are made
- Effect data locality



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Map vs. Reduce

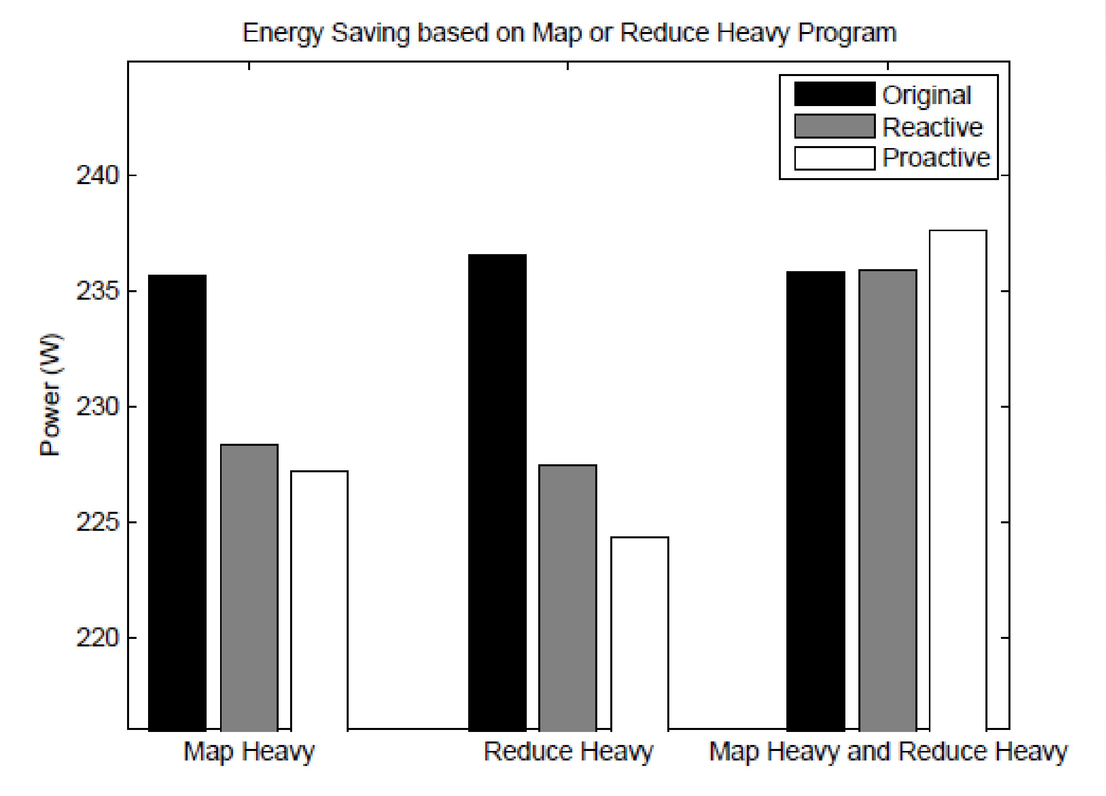
- Map is more I/O intensive usually
- Reduce was usually shorter



AUBURN UNIVERSITY
SAMUEL GINN COLLEGE OF ENGINEERING

Map Heavy vs. Reduce Heavy

- Map Heavy is more I/O intensive
- Map and Reduce Heavy gets no gain



PRE-BUD Model

- Prefetching Energy-Efficient Parallel I/O Systems with buffer Disk

$$E_S(\text{block}(T_{ij})) = E_{WOP} - (E_{WPF} + E_{BUD}).$$

$$\begin{aligned} E_{PF}(P,D) &= E_{R,PF}(P,D) + E_{W,PF}(P,D) \\ &= \sum_{i=1}^{m}\sum_{k=1}^{q}\left(z_{k,i} \cdot P_{A,i} \cdot \left(t_{SK,k,i} + t_{RT,k,i} + \frac{s_{k,i}}{B_{R,i}}\right)\right) \\ &\quad + \sum_{i=1}^{m}\sum_{k=1}^{q}\left(z_{k,i} \cdot P_{A,0} \cdot \left(t_{SK,k,0} + t_{RT,k,0} + \frac{s_{k,i}}{B_{W,0}}\right)\right) \end{aligned}$$

NAP Energy Model

- Find added energy by disks
- Group can either be standby or active
- Read and writes assumed same

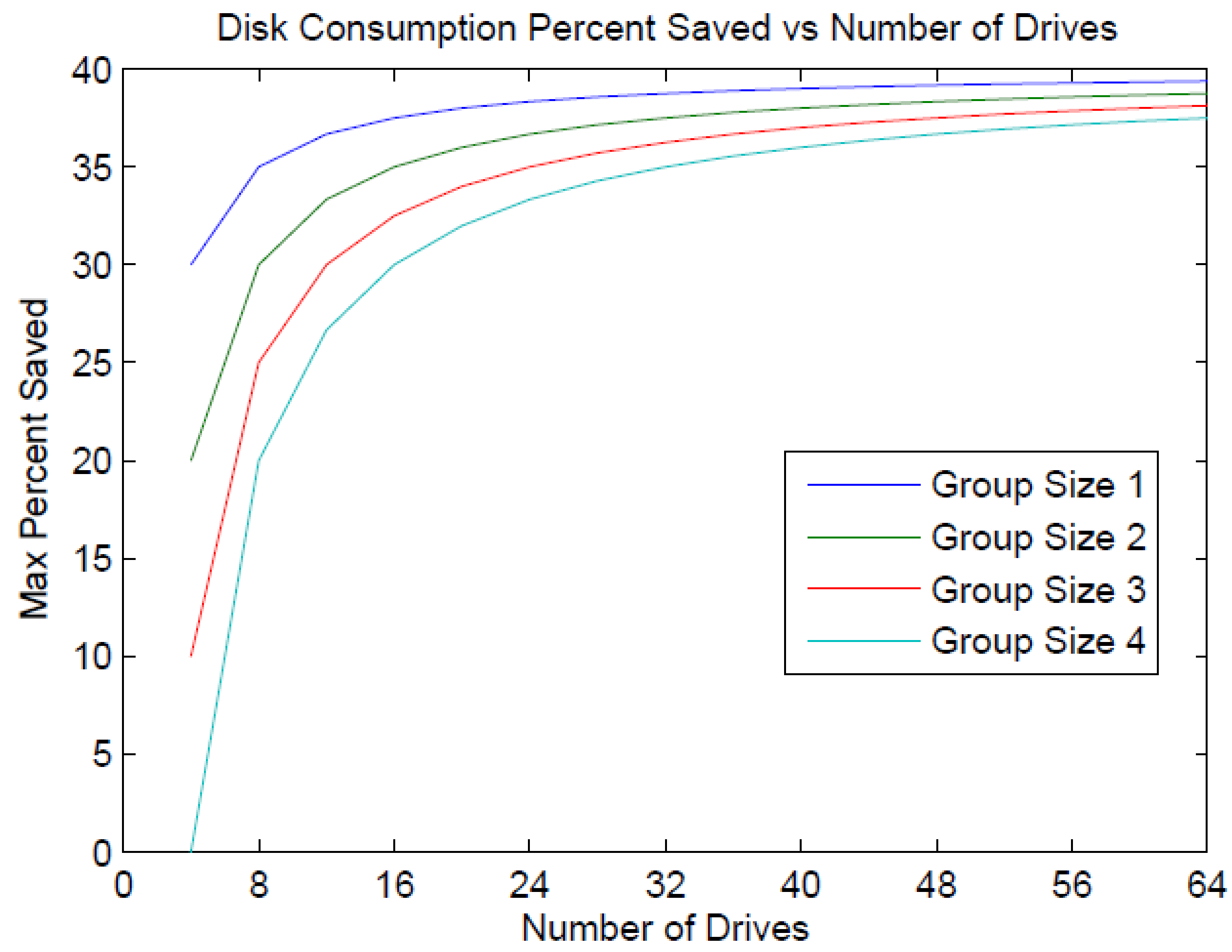
$$E_{total} = E_{server} + E_{disks}$$

$$E_{disks} = \sum_{i=1}^{D/N} \frac{N}{D} E_{group_{i*\frac{D}{N}}} + \sum_{i=1}^{D} \frac{D-N}{D} E_{standby_i} + E_{transitions_i}$$

$$E_{group_n} = \sum_{i=n}^{N+n} E_{active_i}$$

AUBURN UNIVERSITY
Samuel Ginn College of Engineering

Energy Saving Simulation



AUBURN UNIVERSITY

SAMUEL GINN COLLEGE OF ENGINEERING

Summary

- iTad: a simple and practical way to estimate the temperature of a data node
- NAP: an energy-saving technique for disks in Hadoop clusters

Questions