

CS528

Energy Efficient Scheduling in Data Center

A Sahu

Dept of CSE, IIT Guwahati

Energy Efficient System: Design and Management

- **Point to consider**
 - ✓ Energy efficient Infrastructure
 - ✓ Energy Model of Infrastructure
 - Blades/Server Machine CPU, Memory
- 1. Energy Efficient Scheduling**
 - How to manage the Jobs

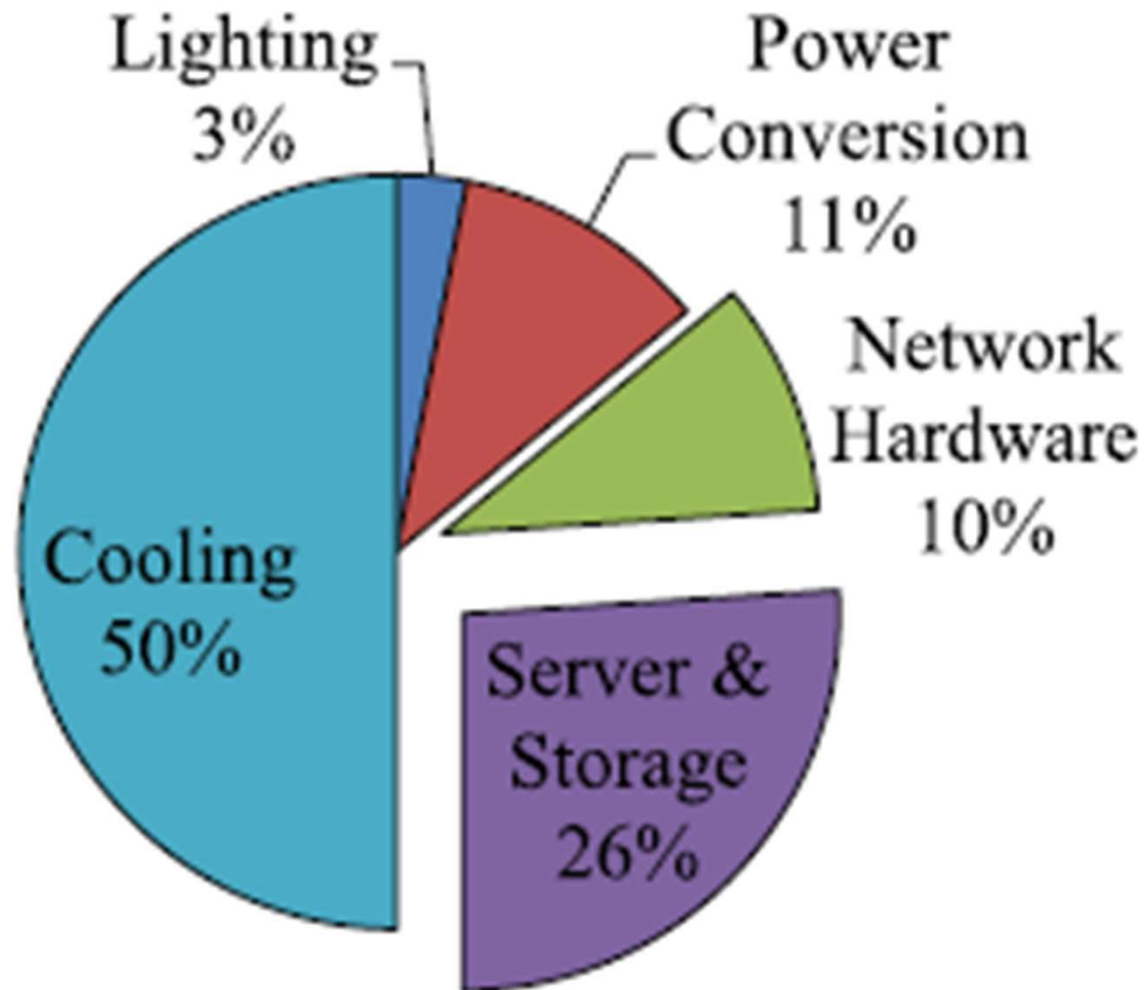
Outline

- ✓ Power/Energy Consumption Model
- Thermal Aware Computing
- Power Aware Scheduling in Cloud
- Migration and Management

Data Center Power Consumption

- Currently it is estimated that servers consume 0.5% of the world's total electricity usage.
 - Closer to 1.2% when data center systems are factored into the equation.
- Server energy demand doubles every 4-6 years.
- This results in large amounts of CO₂ produced by burning fossil fuels.
- What if we could reduce the energy used with minimal performance impact?

Percentage of Power Consumption in DC



Motivation for Green Data Centers

- Economic
 - New data centers run on the Megawatt scale, requiring millions of dollars to operate.
 - Recently institutions are looking for new ways to reduce costs, no more “blank checks.”
 - Many facilities are at their peak operating envelope, and cannot expand without a new power source.

Motivation for Green Data Centers

- Environmental
 - 62% of the U.S. energy sources are fossil fuels. In India 60% is from Coal Plant (NTPC)
 - 2.8 billion tons of CO₂ emitted each year from U.S. power plants.
 - Sustainable energy sources are not ready.
 - Need to reduce energy dependence until a more sustainable energy source is deployed.

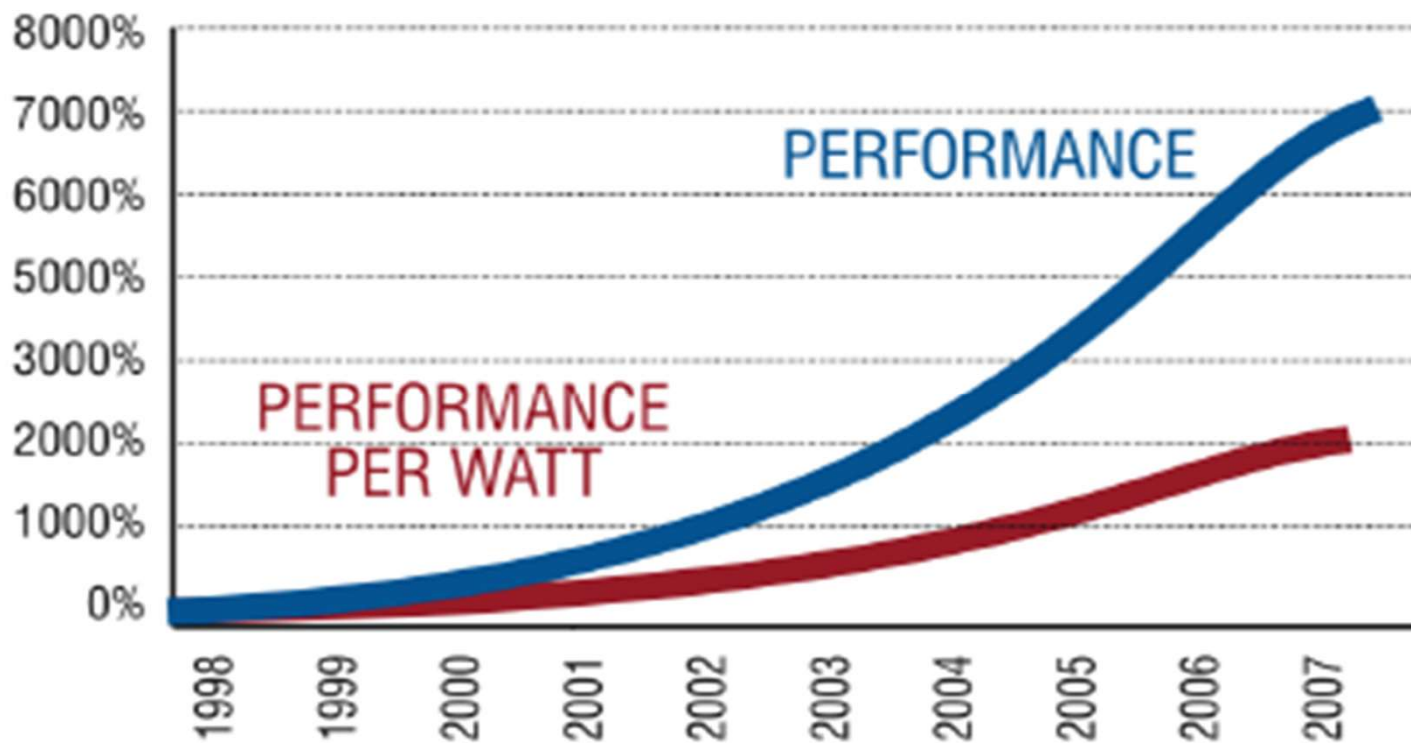
Green Cloud Goal Shift:

“performance” → “energy efficiency”

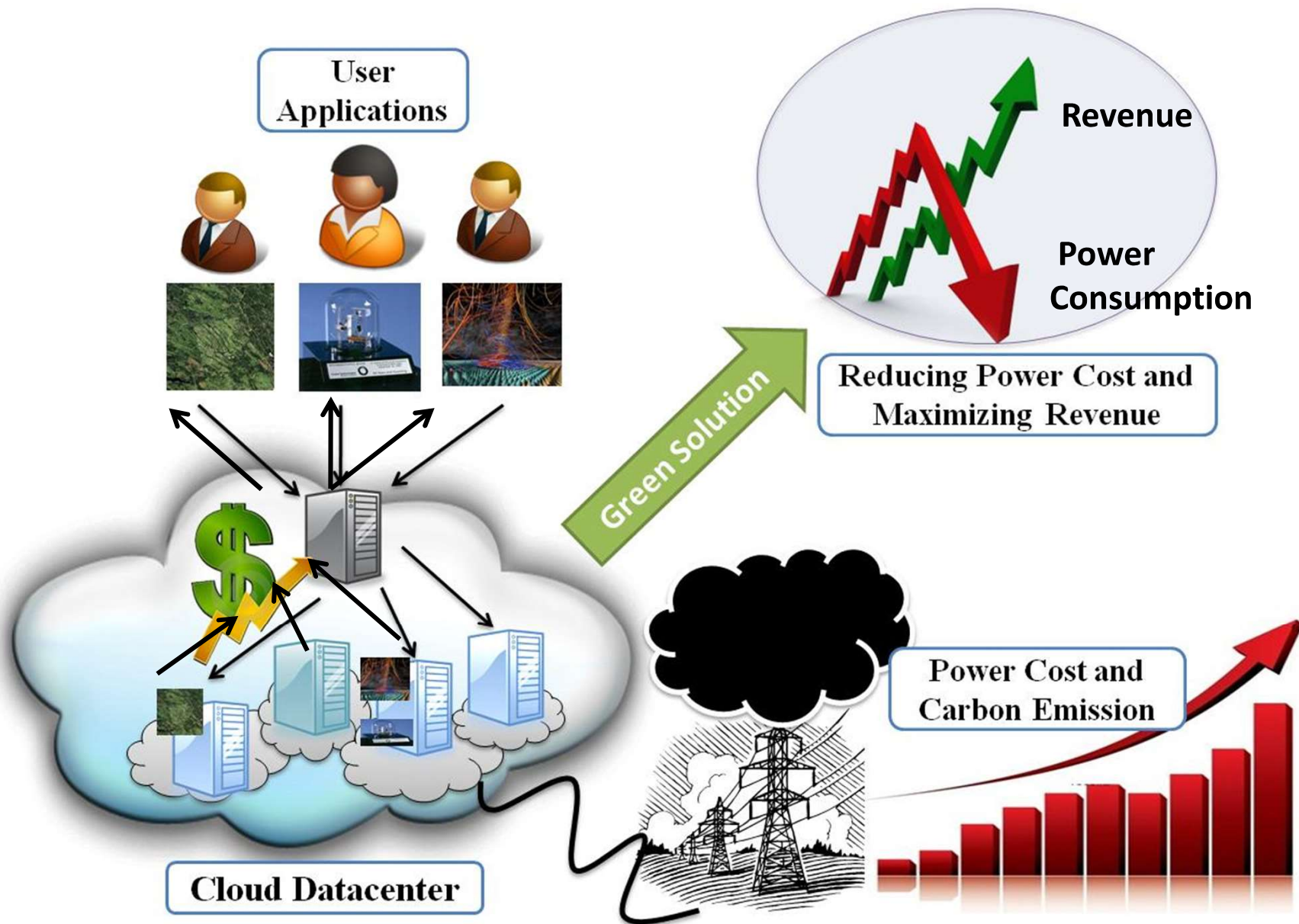
- As **energy costs** are **increasing** while **availability dwindles**
- Need to **shift focus** optimising data center Resource management
 - **From pure performance alone to optimising for energy efficiency**
 - While maintaining high service level performance.
- **Green Cloud computing** model that achieves
 - not only efficient processing and utilisation of computing infrastructure,
 - but also minimise energy consumption.

Green Computing

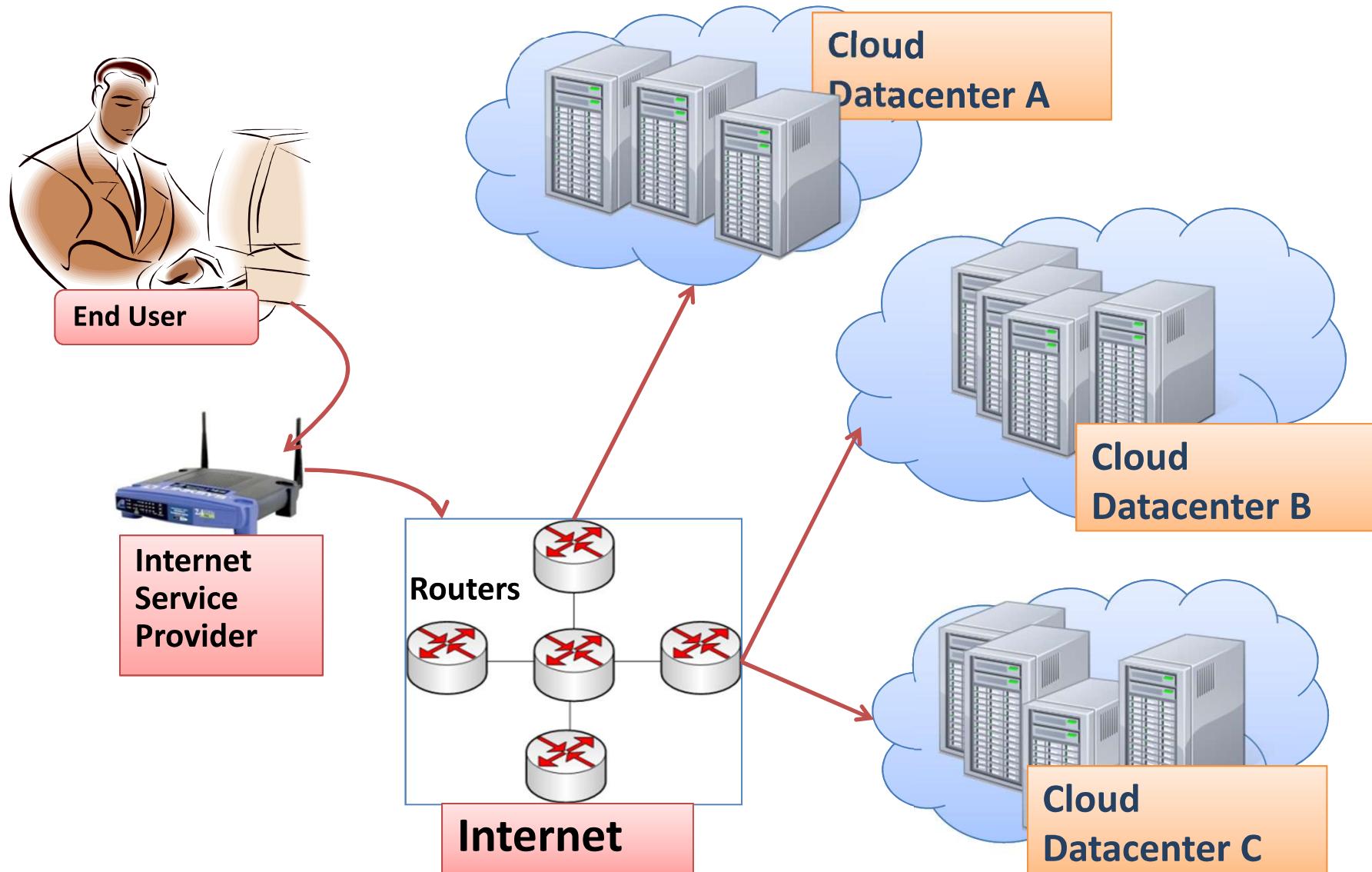
- In the past 15-20 years of supercomputers
 - performance has doubled > 3000 times
 - performance per watt has doubled 300 times
 - performance per square foot has doubled 65 times



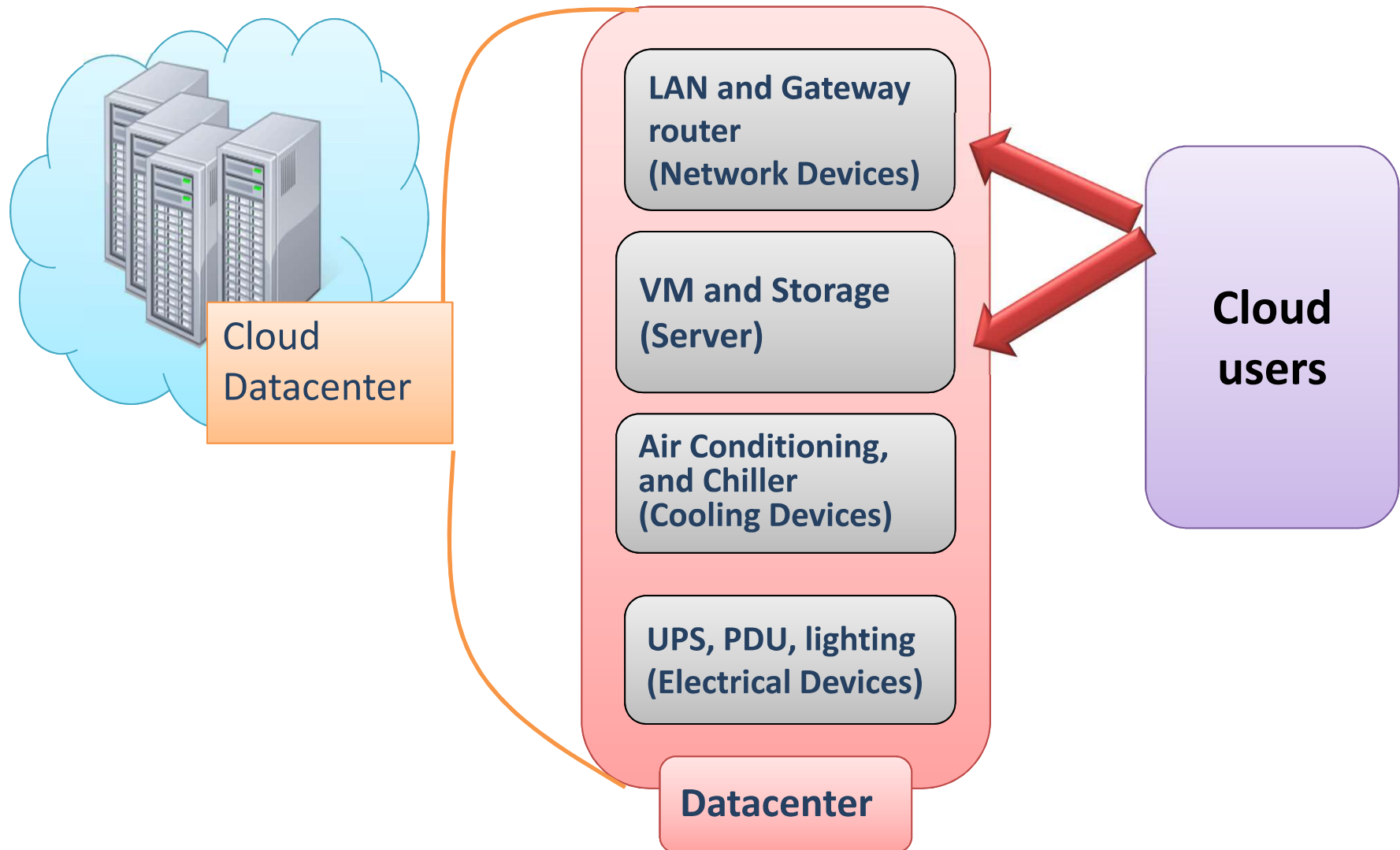
Green Cloud Computing



Cloud Usage Model

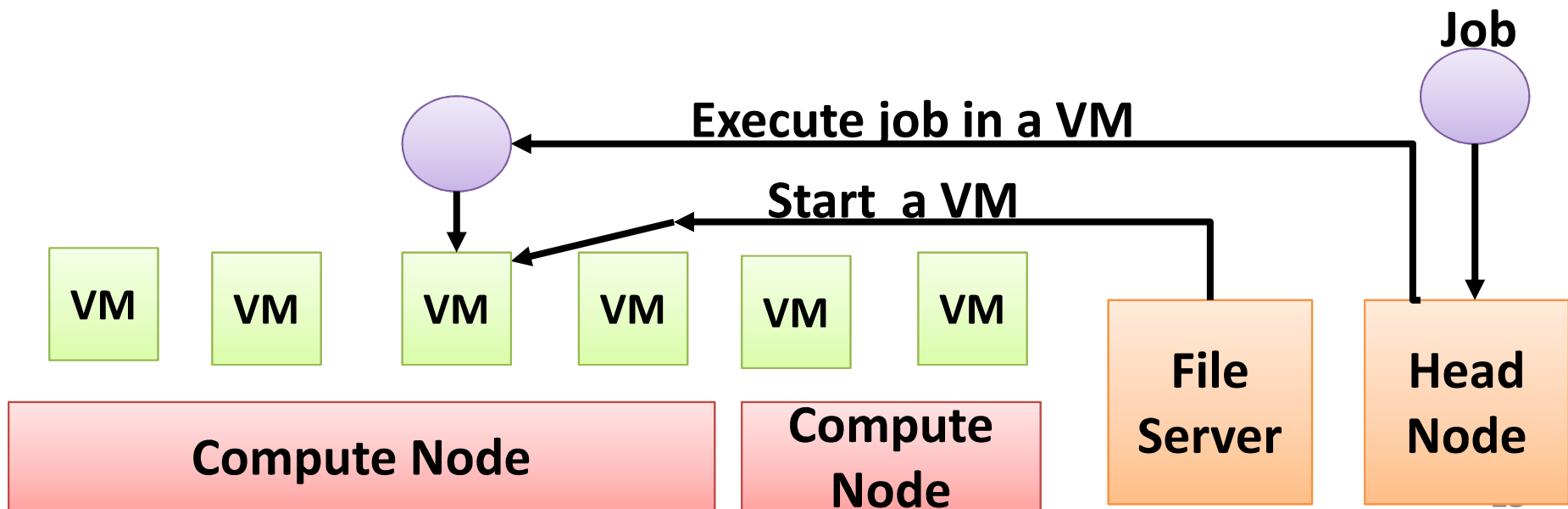


Cloud Usage Model

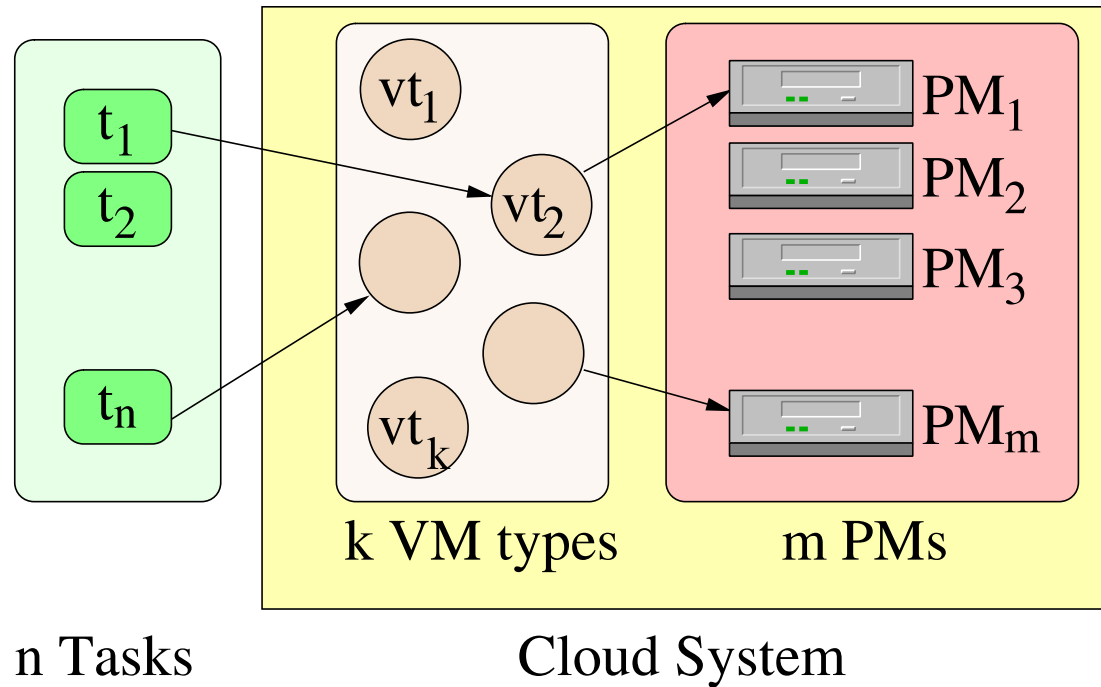


Cloud Computing

- Features of Clouds
 - Scalable, Enhanced Quality of Service (QoS)
 - Specialized and Customized, Cost Effective
 - Simplified User Interface



System model



- Considered IaaS paradigm
- Consists of m homogenous hosts (m is large)

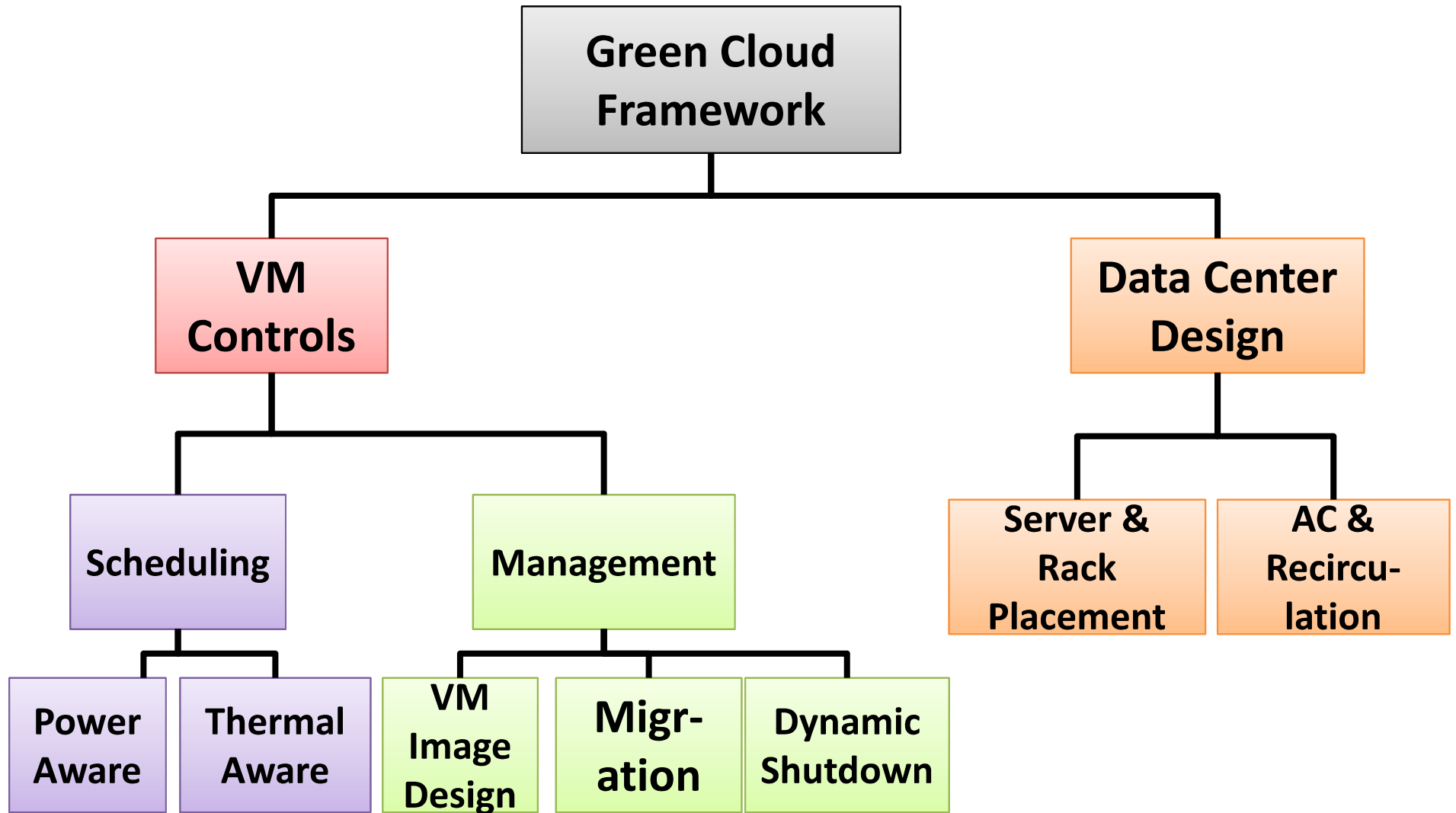
System Model

- Every Task comes with
 - Execution time (e_i), deadline (d_i)
 - CPU requirement (c_i), memory requirement (m_i)
 - and any other
- SLA (Service Level Agreement) Violation (**SLAV**)
 - If the task do not get require amount CPU, memory
 - If competition time extend the deadline
 - Cloud service provide may need to pay penalty

System Model

- Some time SLA comes with
 - Infra oriented: Amount of CPU, memory, bandwidth
 - Service oriented: number of request/task per time, throughput: #web/db request per time
- SLA (Service Level Agreement) Violation (SLAV)
 - If the task do not get require amount CPU, memory
 - If competition time extend the deadline
 - Not able to provide required throughput
 - Cloud service provide may need to pay penalty

Green Cloud Computing Framework



Green Cloud Framework (cont.)

- Goal : Maximize **performance per watt** in Cloud
 - VM Scheduling
 - VM Image Management
 - Data Center Design
- Scheduling
 - Placement within cloud infrastructure
 - Energy use of server equipment
 - datacenter temperature important

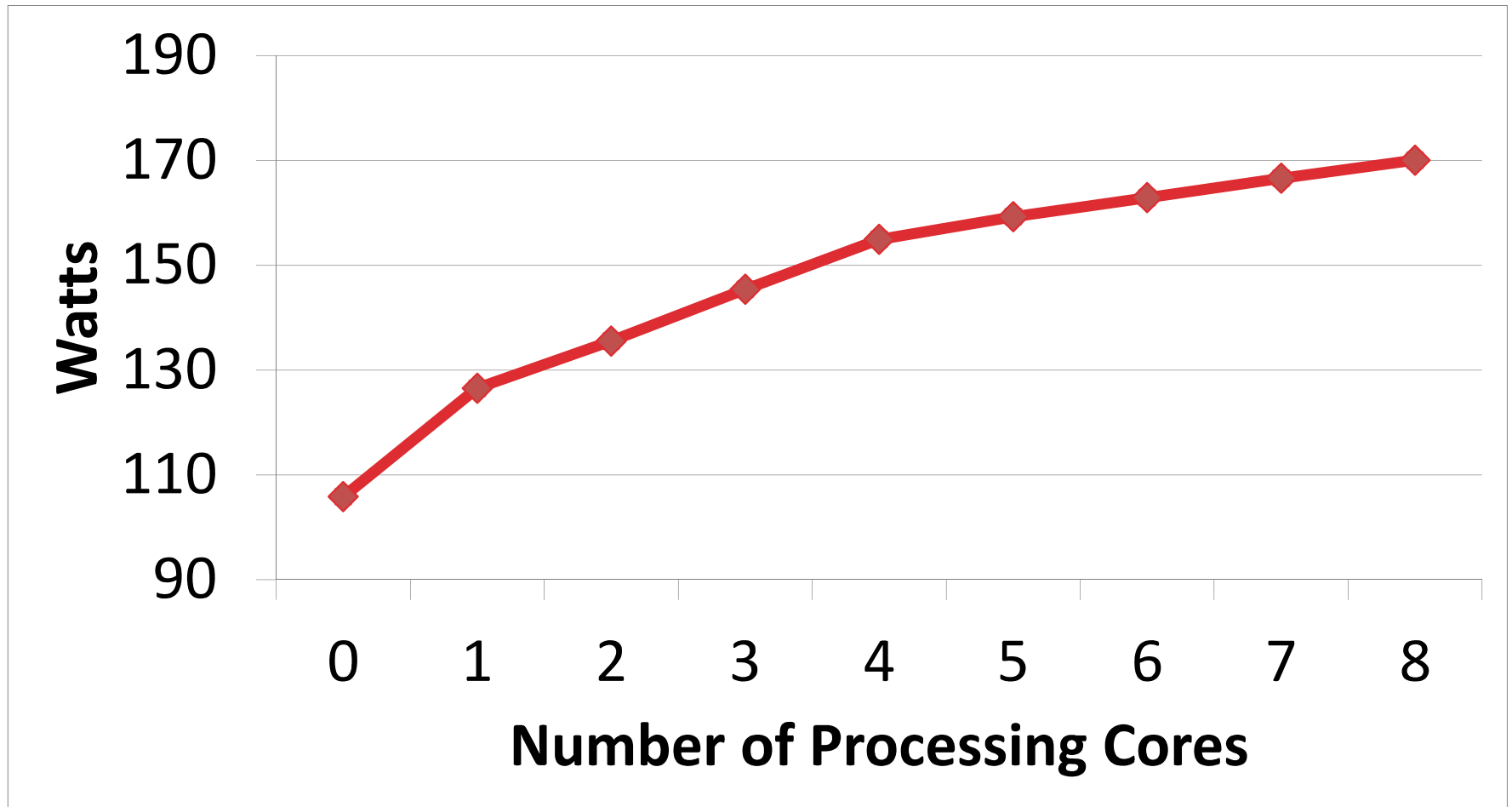
Green Cloud Framework (cont.)

- Image Management
 - Small Size
 - Few unnecessary processes/services
 - Migration
 - Dynamic Shutdown
- Data Center Design
 - More efficient A/C, power supplies
 - Hot and cold aisles
 - Utilizing external cooling

Virtual Machine Scheduling

- Power-Aware Scheduling (PAS)
 - Minimize total power used by servers
 - Power to servers is the larger cost
- Thermal-Aware Scheduling (TAS)
 - Minimize overall temperature
 - Reduces energy used for cooling

Virtual Machine Scheduling

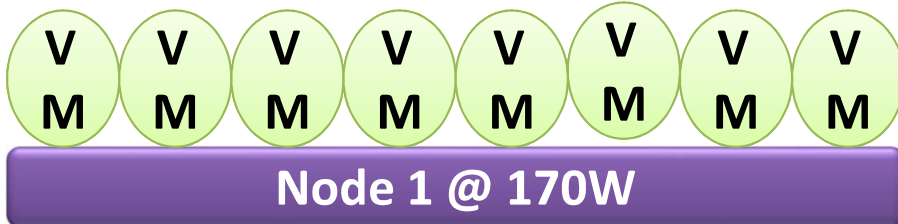


Power consumption curve on an Intel Core i7 920 Server
(4 cores, 8 virtual cores with Hyperthreading)

Power Aware (PA) Computing

- Objective of PA computing/communications is
 - To improve power management and consumption
 - Using the awareness of power consumption of devices.
- Power consumption is most important considerations
 - In mobile devices due to limitation battery life.

485 Watts vs. 552 Watts



$$105 * 3 + 170 = 485$$



vs.

