Bits of Architecture

Introduction to Performance

What Does Better Performance Mean?

Which is Better?

Sports Car

- Fast
- Limited seating
- Poor Fuel Efficiency (per-person)

Sedan

- Moderate Speed
- Moderate Seating
- Decent Fuel Efficiency (per-person)

Tour Bus

- Slow
- Lots of Seating
- Good Fuel Efficiency (per-person)

Takeaway: The best choice depends on our needs (and is rarely clear-cut)

Performance Metrics

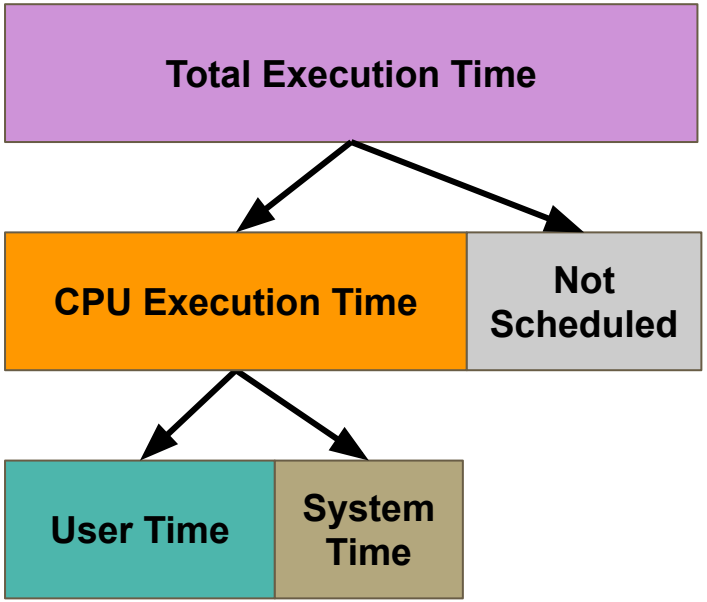
Performance Metrics

- **Execution Time**
  - Total time to complete a task
- **Throughput**
  - Also known as **Bandwidth**
  - Number of tasks completed per unit time

Time Spent in the CPU

How We Measure CPU Time

- **CPU Execution Time**
  - Time spent actively working on a task
- **User CPU Time**
  - Time spent in the program
- **System CPU Time**
  - Time spent in the OS on behalf of the program



Performance Is All About Ratios

The Performance Ratio

How do we define performance?

$$Performance_x = \frac{1}{ExecutionTime_x}$$

How do we compare the performance of 2 systems?

$$\frac{Performance_x}{Performance_y} = \frac{ExecutionTime_y}{ExecutionTime_x}$$

Computer X runs an app in 10s
Computer Y runs an app in 15s
How much faster is X compared to Y?

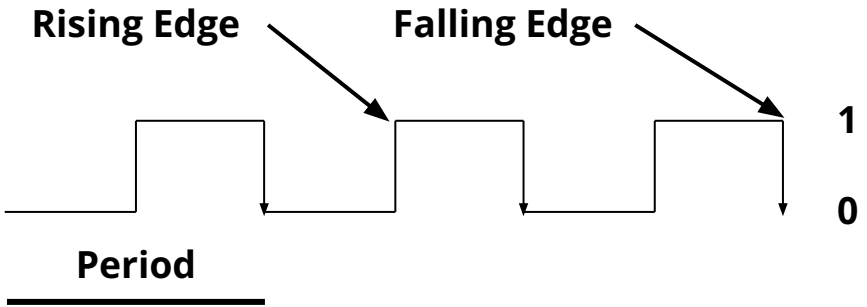
$$\frac{15s}{10s} = 1.5$$

The Clock

The Clock

- **Clock**
  - Generates an oscillating signal
  - Drives execution
- **Clock Cycle**
  - One period of our clock signal
- **Frequency (Hz)**
  - Occurrences per unit time
  - 20kHz = 20,000 clock periods / s

- We often talk about instructions in terms of how many cycles they take to execute



Instruction Performance

Programs = Instructions

- **Clock Cycles per Instruction (CPI)**
  - Average number of clock cycles taken per instruction
  - The inverse (**IPC**) is also incredibly common

$$ExecutionTime_{cpu} = CPI \times InstructionCount \times ClockPeriod$$

- To predict performance, we often look at the cost of each instruction (or class of instruction)

$$TotalCycles = \sum_{i=1}^{n} (CPI_i \times C_i)$$

Caveats…

It Depends...

- Does a higher CPI on one machine mean it's slower than another machine?
- Does a machine having a higher CPI on one application mean it's worse than another machine?
- Do instructions always take a fixed length of time?

- **Execution time is our ground truth!**