

### Web-Scale Problems?

- Don't hold your breath:
  - Biocomputing
  - Nanocomputing
  - Quantum computing
  - ...
- It all boils down to...
  - Divide-and-conquer
  - Throwing more hardware at the problem

Simple to understand... a lifetime to master...

WiSe Lab @ WMU  
www.cs.wmich.edu/wise
Cloud Computing, July 2015
90

---

---

---

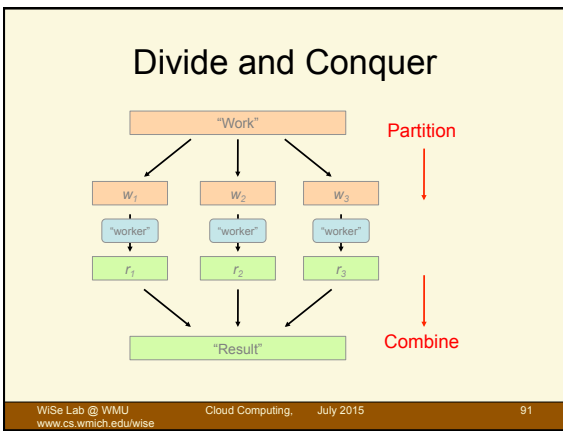
---

---

---

---

---




---

---

---

---

---

---

---

---

### Different Workers

- Different threads in the same core
- Different cores in the same CPU
- Different CPUs in a multi-processor system
- Different machines in a distributed system

WiSe Lab @ WMU  
www.cs.wmich.edu/wise
Cloud Computing, July 2015
92

---

---

---

---

---

---

---

---

### Choices, Choices, Choices

- Commodity vs. “exotic” hardware
- Number of machines vs. processor vs. cores
- Bandwidth of memory vs. disk vs. network
- Different programming models

WiSe Lab @ WMU      Cloud Computing, July 2015      93  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

### Flynn's Taxonomy

		Instructions	
		Single (SI)	Multiple (MI)
Data	Single (SD)	<b>SISD</b> Single-threaded process	<b>MISD</b> Pipeline architecture
	Multiple (MD)	<b>SIMD</b> Vector Processing	<b>MIMD</b> Multi-threaded Programming

WiSe Lab @ WMU      Cloud Computing, July 2015      94  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

### SISD

The diagram shows a central blue box labeled 'Processor'. An arrow labeled 'Instructions' points up to the processor. To the left, three orange boxes labeled 'D' are connected by an arrow pointing into the processor. To the right, three green boxes labeled 'D' are connected by an arrow pointing away from the processor.

WiSe Lab @ WMU      Cloud Computing, July 2015      95  
www.cs.wmich.edu/wise

---

---

---

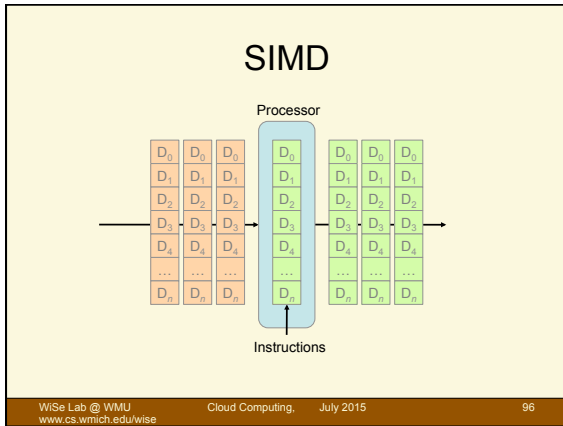
---

---

---

---

---



---

---

---

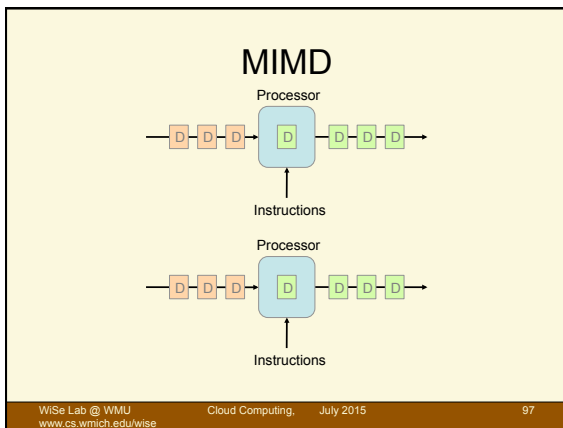
---

---

---

---

---



---

---

---

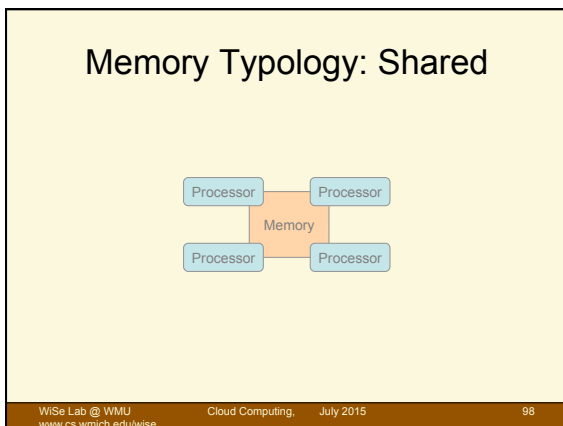
---

---

---

---

---



---

---

---

---

---

---

---

---

### Memory Typology: Distributed

The diagram illustrates a distributed memory architecture. It consists of four nodes arranged in a 2x2 grid. Each node contains a 'Processor' (blue box) and 'Memory' (orange box) connected by a vertical line. A horizontal double-headed arrow labeled 'Network' connects the nodes across the top and bottom rows.

WiSe Lab @ WMU  
www.cs.wmich.edu/wise Cloud Computing, July 2015 99

---

---

---

---

---

---

---

---

### Memory Typology: Hybrid

The diagram illustrates a hybrid memory architecture. It consists of four nodes arranged in a 2x2 grid. Each node contains two 'Processor' (blue boxes) and one 'Memory' (orange box) connected by a vertical line. A horizontal double-headed arrow labeled 'Network' connects the nodes across the top and bottom rows.

WiSe Lab @ WMU  
www.cs.wmich.edu/wise Cloud Computing, July 2015 100

---

---

---

---

---

---

---

---

### Parallelization Problems

- How do we assign work units to workers?
- What if we have more work units than workers?
- What if workers need to share partial results?
- How do we aggregate partial results?
- How do we know all the workers have finished?
- What if workers die?

What is the common theme of all of these problems?

WiSe Lab @ WMU  
www.cs.wmich.edu/wise Cloud Computing, July 2015 101

---

---

---

---

---

---

---

---

**General Theme?**

- Parallelization problems arise from:
  - Communication between workers
  - Access to shared resources (e.g., data)
- Thus, we need a synchronization system!
- This is tricky:
  - Finding bugs is hard
  - Solving bugs is even harder

WiSe Lab @ WMU      Cloud Computing, July 2015      102  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

**Managing Multiple Workers**

- Difficult because
  - (Often) don't know the order in which workers run
  - (Often) don't know where the workers are running
  - (Often) don't know when workers interrupt each other
- Thus, we need:
  - Semaphores (lock, unlock)
  - Conditional variables (wait, notify, broadcast)
  - Barriers
- Still, lots of problems:
  - Deadlock, livelock, race conditions, ...
- Moral of the story: be careful!
  - Even trickier if the workers are on different machines

WiSe Lab @ WMU      Cloud Computing, July 2015      103  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---

**Patterns for Parallelism**

- Parallel computing has been around for decades
- Here are some “design patterns” ...

WiSe Lab @ WMU      Cloud Computing, July 2015      104  
www.cs.wmich.edu/wise

---

---

---

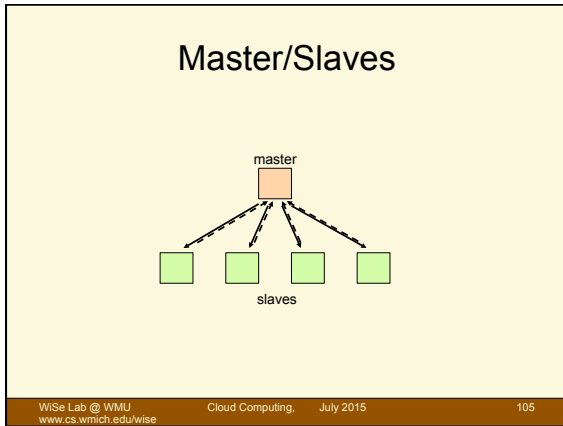
---

---

---

---

---



---

---

---

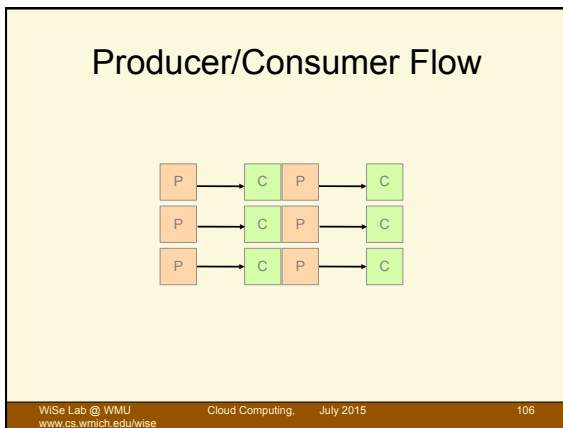
---

---

---

---

---



---

---

---

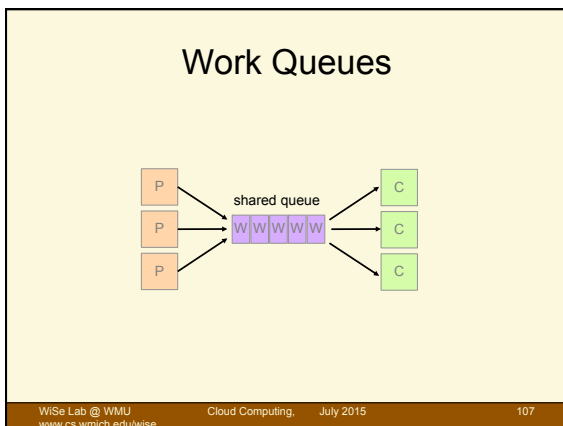
---

---

---

---

---



---

---

---

---

---

---

---

---

### Rubber Meets Road

- From patterns to implementation:
  - pthreads, OpenMP for multi-threaded programming
  - MPI for clustering computing
  - ...
- The reality:
  - Lots of one-off solutions, custom code
  - Write you own dedicated library, then program with it
  - Burden on the programmer to explicitly manage everything
- MapReduce to the rescue!
  - (for next time)

WiSe Lab @ WMU      Cloud Computing, July 2015      108  
www.cs.wmich.edu/wise

---

---

---

---

---

---

---

---