INTRODUCTION TO MPI AND HIGH PERFORMANCE COMPUTING

Prashant Piprotar

IIT Kanpur

15th May, 2020

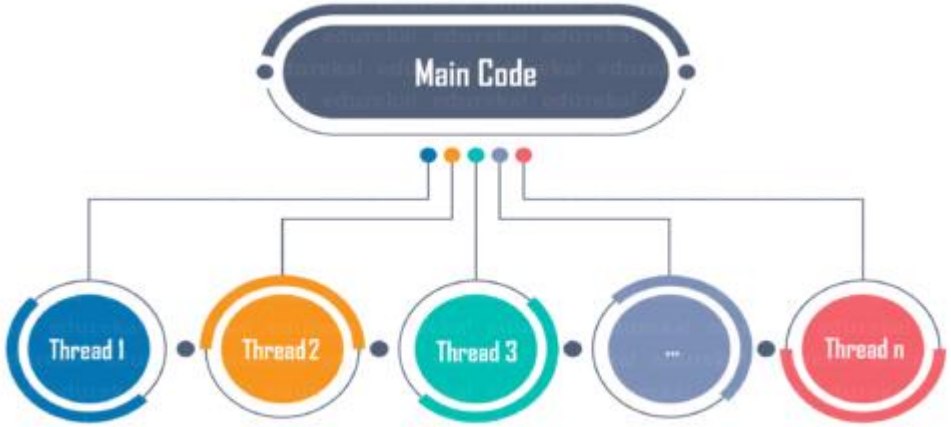
Story Time …

- Dr. Ritesh Patel got us a simulation task from one of his student at IIT.
- **Problem**: Task was taking weeks to complete.
- **Simple Solution**: Multithreading. Right?

```
for(i = 0;i<4;i++){      -> Break down this loop to make it 4 parts.
        for( ...  ){
                for( ... ){ }
        }
}
```

Calculations:

- Original task Duration: 8 Days (assumed)
- Fastest PC (at that time): Quad Core 4th gen i7.
- New Duration: About 2 Days



Multithreading to Clustering

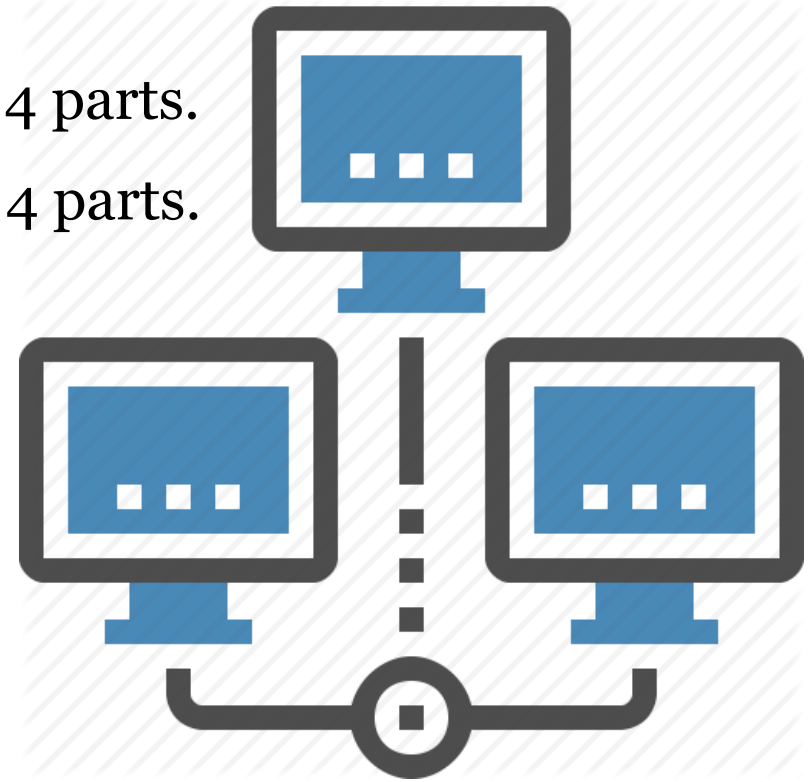
- **Final Solution:** Multi threading + Multiple hosts
- **No of Hosts:** 4

Some Calculations:

```
for(i = 0;i<4;i++){                 -> Break down this loop into 4 parts.
        for(j=0;j<4;j++ ){          -> Break down this loop into 4 parts.
                for( ... ){ // Tasks }
        }
}
```

4 PC * 4 Cores = 16 Cores

- **Completion Time:** Just few hours.



High Performance Computing

What is High Performance Computing?

- For 1st year Undergrad:

1. Computing on High Performance Computers.
2. Solving problems at very large scales.
3. Simulations / Analysis that takes days to complete.

- Who uses HPC?

1. Weather and climate research
2. Bioscience and energy sector
3. Automotive, aerospace, oil and gas exploration
4. Financial simulation, etc.

- Similar Concepts :

1. Supercomputing
2. Parallel Computing

Why do we need them?

- Its Simple
- **To do things Faster.**

What does HPC include?

- Two main Hardware elements:

1. Compute Elements (for Computation) : CPU, GPU, etc
2. Network Elements (for Communication) : Infiniband, Ethernet, etc

- Two main Software elements:
- Programming Models : MPI, Shared memory
- (IMP) Your Application

What makes it HPC?

All the above things but at Larger Scale.

HPC Computer Architectures

Vector Computers (VC) - proprietary system

- – Provided the breakthrough needed for the emergence of computational science, but they were only a partial answer

Massively Parallel Processors (MPP) - proprietary systems

- – High cost and a low performance/price ratio.

Symmetric Multiprocessors (SMP)

- – Suffers from scalability

Distributed Systems

- – Difficult to use and hard to extract parallel performance

[Widely Used] Clusters – commodity and highly popular

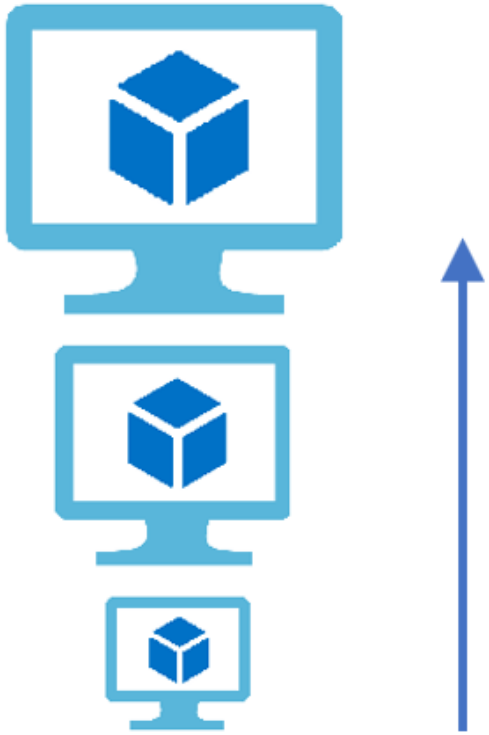
- – High Performance Computing - Commodity Supercomputing
- – High Availability Computing - Mission Critical Applications

Horizontal vs Vertical Scaling

www.abhijitkakade.com

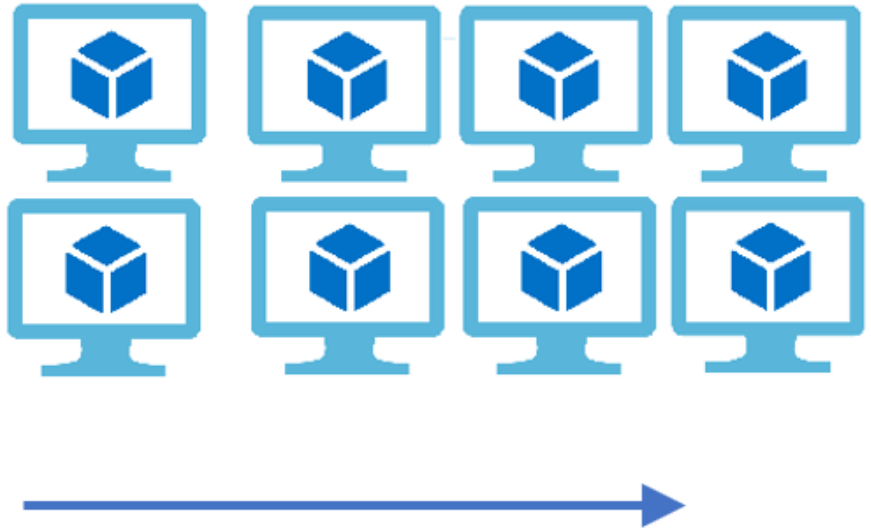
Vertical Scaling

(Increase size of instance (RAM , CPU etc.))

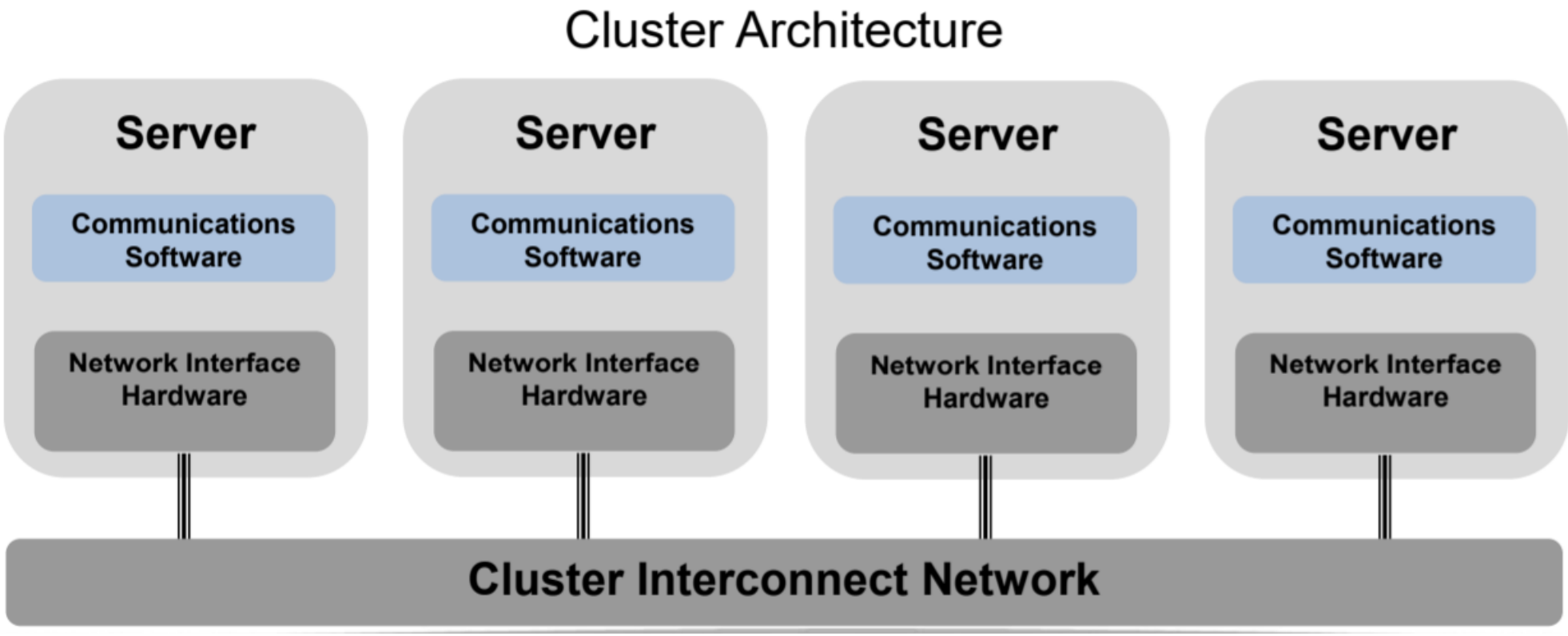


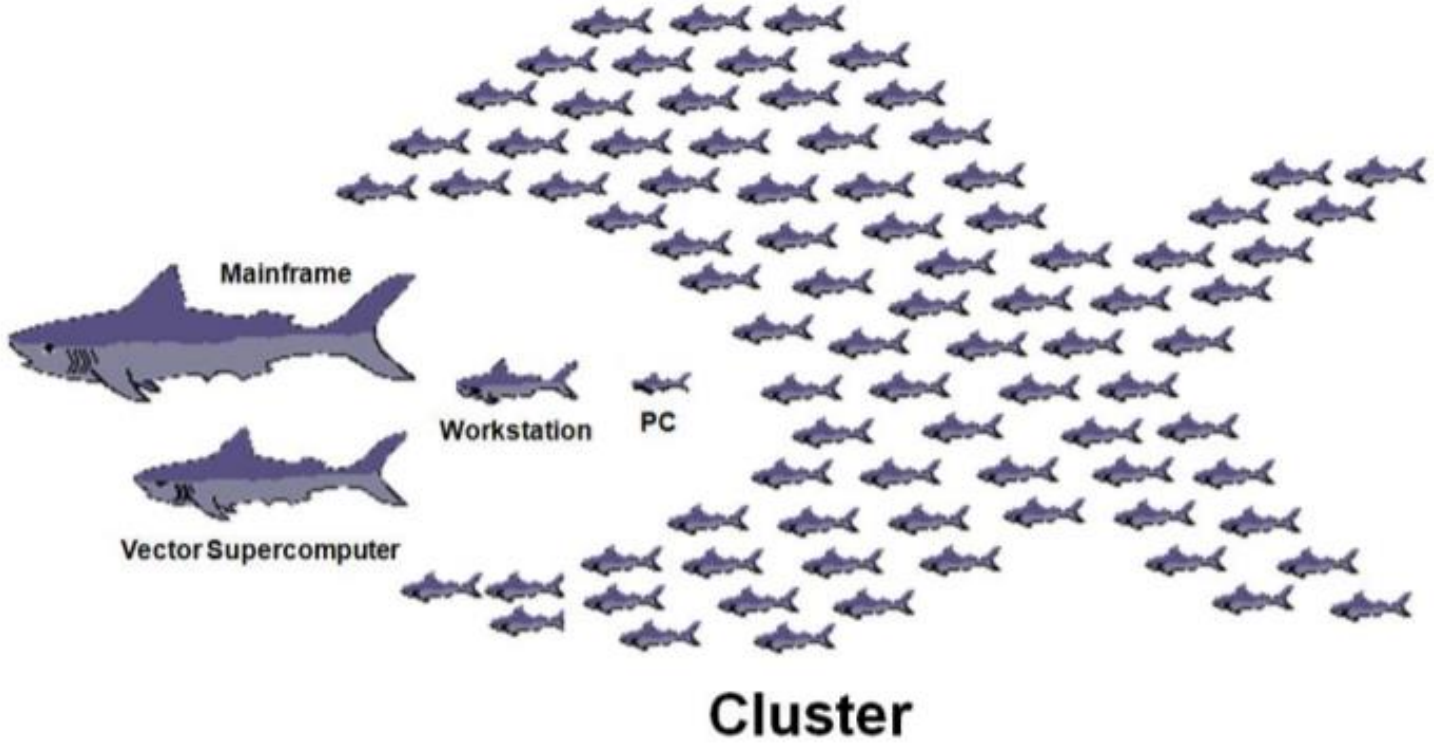
Horizontal Scaling

(Add more instances)



HPC Clusters – Affordable, Efficient and Scalable HPC Solution

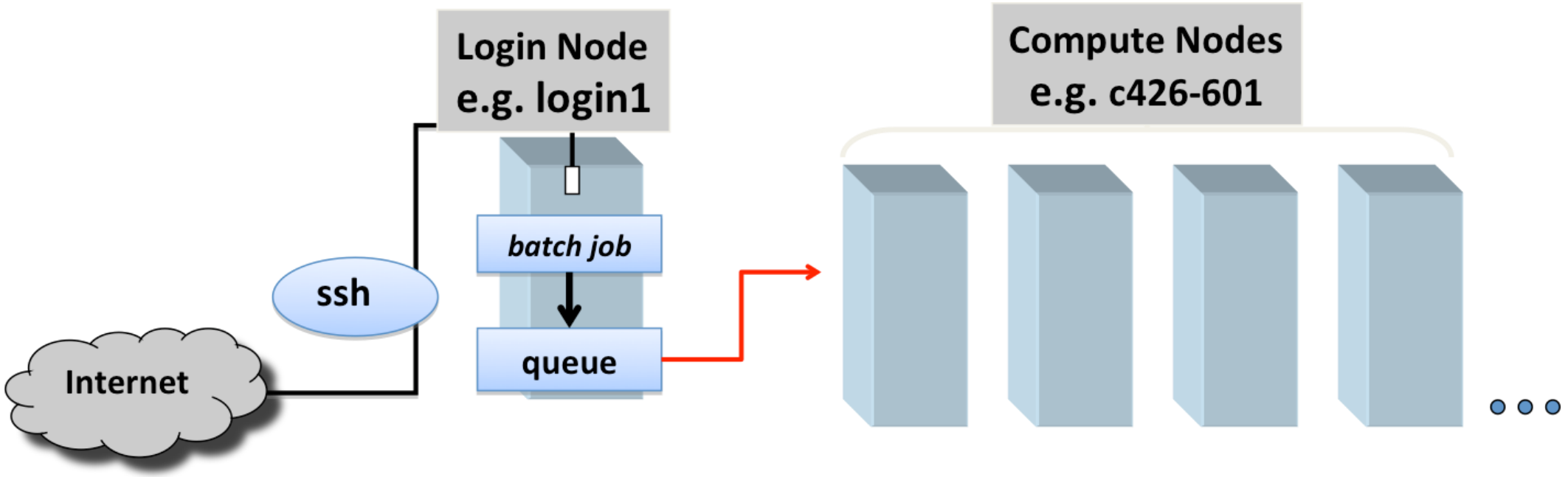




COMPUTER FOOD CHAIN

Parallel Computing on a Large Number of Servers is More Efficient than using Specialized Systems

HPC Queue System



INTRODUCTION TO MPI

MPI- Message Passing Interface

- A standard for Message passing library.
- Efficient, Portable, Scalable, Vendor Independent.
- Supports: Distributed, Shared, Hybrid Memory architectures.
- Some Implementations:

1. MPICH
2. OpenMP
3. Intel MPI
4. Microsoft MPI, etc.

Demo 0: Hello World

Demo 1: Send Receive

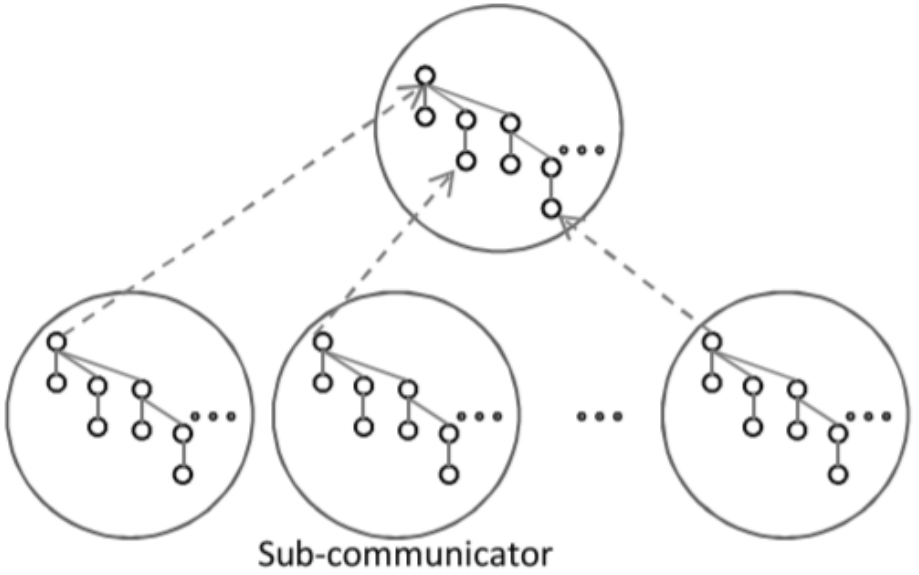
Demo 2: Broadcast

Demo 3: A brief about my CloudMPI

Approach

As Inspired by : Network Performance Aware MPI Collective Communication Operations in the Cloud (Yifan Gong)

- Using existing MPI collectives
- Over hierarchical tree (sub communicators)



Conclusion

- Cluster is right now “the best” architecture for building HPC.
- Queue systems helps to increase the utilization.
- MPI reduces the efforts to operate HPC.
- You can make group projects and running on cluster made by lab PCs.

Q & A

Thank You

If you have any queries, mail it to prashplus@gmail.com

Part of the content has been referred from: https://www.hpcadvisorycouncil.com/