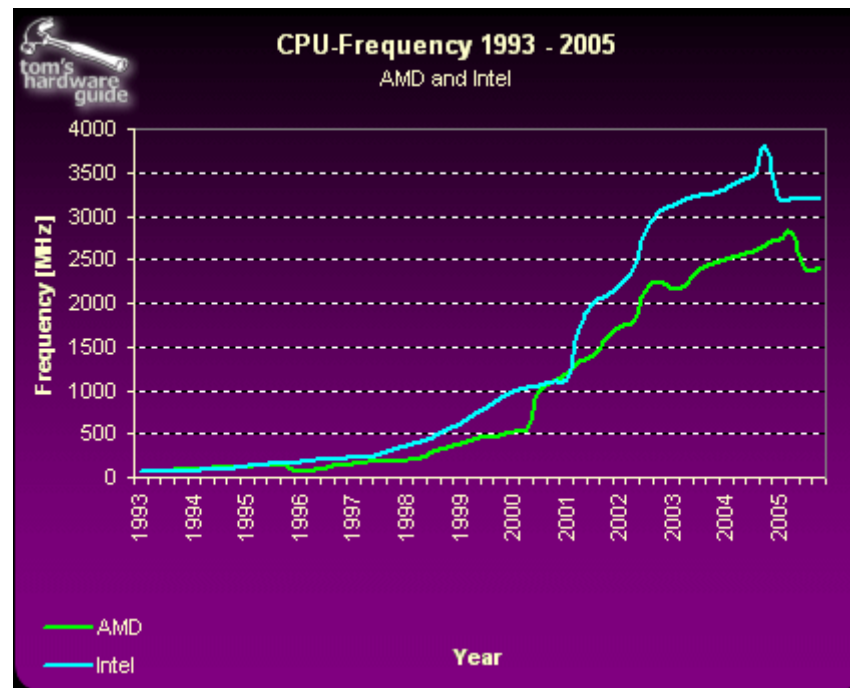# Something about Distributed & Parallel Computing

jingmi@baidu.com

2008-09-20～2008-09-22

Computer Speedup

Problems

- Rendering multiple frames of high-quality animation (Hollywood)

- Indexing the web (Search Engine : Semi-Structured Data)

- Speeding up content delivery (Akamai)

- Data Storage / Query (ga, tongji.baidu.com : Structured Data)

Example

- GFS / Bigtable
- Amazon EC2 (Amazon Elastic Compute Cloud)
- Amazon Dynamo
- Google App Engine
- Oracle RAC
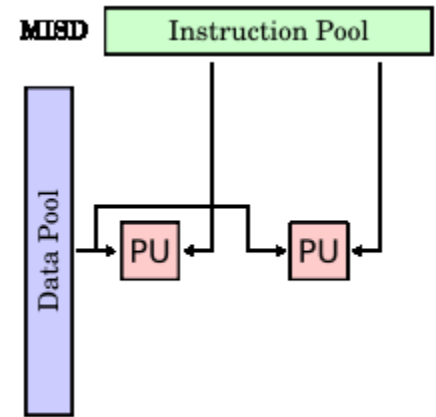- Dryad / Boxwood
- Beowulf

Conception

- Distribution
- Parallel
- Concurrence
- HPC / HA
- CMP / SMP
- NUMA / UMA / ccNUMA
- Scale up / Scale out

Flynn's Taxonomy of Computers (CSArch)

SISD	MISD
SIMD	MIMD

MISD

- Multiple Instruction, Single Data
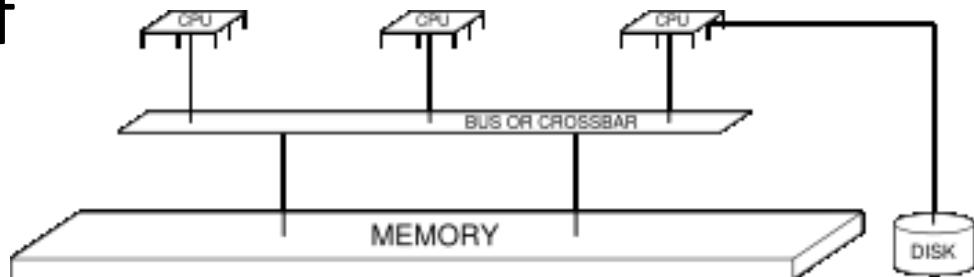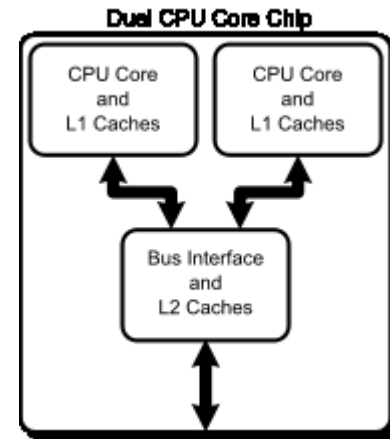- Pipeline
- Super Pipeline / Super Scalar

Relevant to Pipeline

- Example: Tracker 4.0 Prestat
- Bad case: 1 GHz Itanium ≈ 2.66 GHz Xeon / Opteron
- Indicator: throughput、 efficiency、 deep、 speedup
- High pipelining leads to increase of latency.
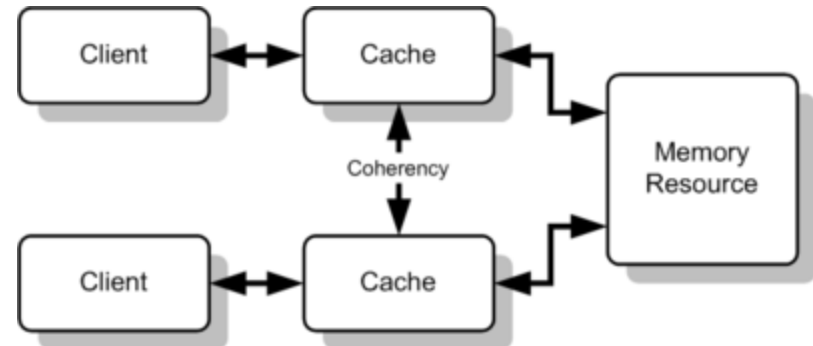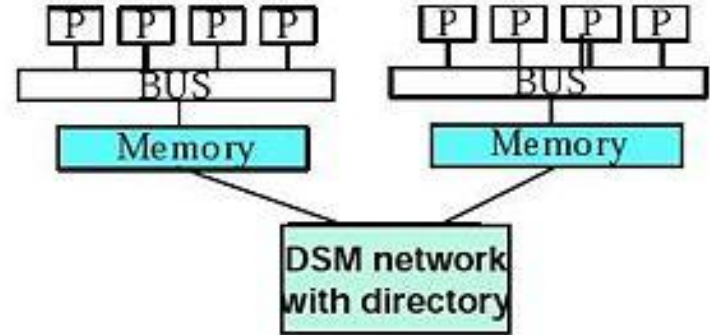- Stall / Predict / Delay / Locality of reference

Mutil-Core

- CMP
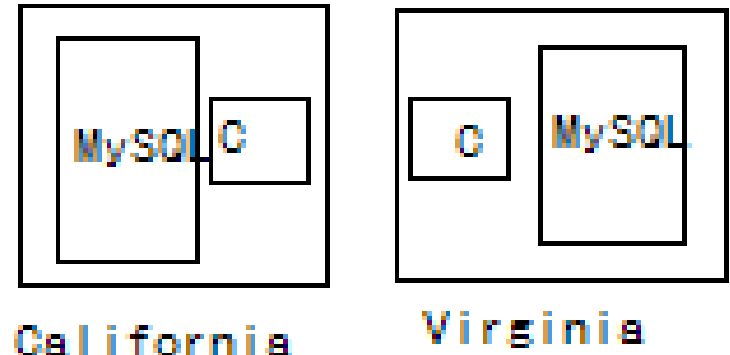
- SMP

- ASMP -- Beowulf

Memory Access

- UMA
- NUMA
- ccNUMA

Cache Consistency

1. Update master 'a' -> 'b'
2. Remove both cache
3. Read slave cache
4. Read slave MySQL(Old info)
5. Insert old info to slave cache
6. Sync slave MySQL
7. Read slave cache (OOPS!)

California

Virginia

http://www.25hoursaday.com/weblog/2008/08/21/HowFacebookKeepsMemcachedConsistentAcrossGeoDistributedDataCenters.aspx

Share

- Share Memory
- Share Storage
- Share Nothing

Data Independence

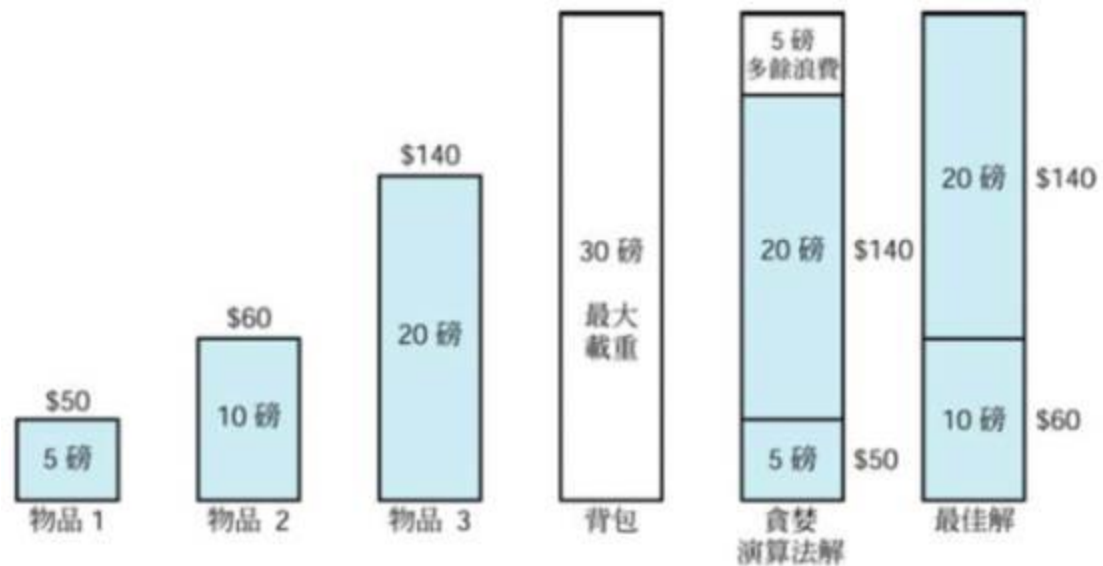- Data Abstraction
- Code - Data

Fp (Why fp matters?)

- No Side Effect - Lazy Evaluation
- High Order Function
- Closure
- MapReduce

Parallelism

- Task parallelism (also known as function parallelism and control parallelism)

- Data parallelism (also known as loop-level parallelism).

- Instruction-level parallelism (ILP).

- Bit-level parallelism: pattern matching.

Task Scheduler

- Greedy Algorithms
- NPC
- AI

CS

- AI
- Programming Language
- System (OSDI, SOSP, NSDI, ISCA, HPCA, MICRO)
- Theory

Reference & Books

- *SICP*
- *CSAPP*
- *Parallel Computer Architecture: A Hardware/Software Approach*
- *Principles and Practices of Interconnection Networks*
- *The Hardware/Software Interface*
- *Parallel Computer Architecture - A Hardware/Software Approach*

History Always Repeat Itself!!!

Thanks !