PDC

Parallel Computing
aik computer k jab multiple processor kam kr rahe hote

Distributed computing
multiple computers jab sath kisi task per kam kr rahe hote

PARALLEL COMPUTING
Performing task across multiple processor

Types
① Bit-level parallelism
② Instruction level parallelism
③ Task level parallelism

DISTRIBUTED COMPUTING
Connecting computers via a network to act together as a powerful unit

Advantages
- Scalability
- redundancy

ایک system خراب بھی ہو جائے تو کام ہوتا رہتا ہے کیونکہ سب PCs مل کر کام کرتے۔ جب کہ parallel میں کام رک جاتا ہے

Memory
Parallel => shared memory, processor

Synchronization speed پر ساتھ ساتھ کام کرنے کی
Parallel => single master clock for all processors (CPU synchronization)
distributed => uses synchronization algo.

Parallel → less cost, slow
Distributed → expensive, management required, fast

⟷

Parallelism → multiple works ایک ساتھ کرنا۔
Speed↑, task ↑
cost ↑
- Task (set of instructions)
- Data (raw form of information)

Bit-level → جو ہمارے system پر ہوتی

Task Parallelism
song + word

Characteristics
- independent tasks
- asynchronous — دونوں کا آپس میں کوئی updation (کام) نہیں
- Heterogeneous work load — کسی task کا load نہیں اور کسی کا کم
- Communication (independent) — ہو سکتی لیکن بہت کم ہوتی۔

Advantages and Challenges
- efficient utilization of multi-core processors
- run concurrently — ایک ساتھ
- can handle heterogeneous workload

Challenges
- managing inter task and synchronization
- task scheduling complex — یہ کام ہم نہیں کرا سکتے
when ~~tasks~~ depend on each other

~~→ task →~~

task scheduling is good when tasks are independent

Thread => subprocess
Multiple thread => process is divided into multiple pieces and execute after combing

Threading
Multiple threading
Task scheduling
Clock related/instruct

```
void main()
{ int a
  cin >> a
void getdata()
{ body }
getdata()
}
```

Data Parallelism
focuses on distributing data and performing same operation

male female → same operation (calculate gpa)

Distributed → job dusre PC's ko job kr kee thee
Shared → aik he ko hum share kite apas mai.

Challenges
Data depend ki raha parallelism nahi hoga
Synchronization nahi hoga

Hybrid => jab dono mil kr kaam krte

Conclusion
Data => large data sets
Task => different tasks

(P2)

LAB

PDC

<u>Principles</u>

Decomposition => problem into small problems if possible

Concurrency => multiple instructions at same time.

Synchronization => same time

Scalability => data increase => problem solve hojaye

NxN

2 matrix

$$\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 5 & 7 \\ 6 & 8 \end{bmatrix}$$

threads, process
↳ library.

$$\begin{bmatrix} 1\times5 + 3\times6 & 1\times7 + 3\times8 \\ 2\times5 + 4\times6 & 2\times7 + 4\times8 \end{bmatrix}$$

① NxN
② random.
③ dot( ) ← numpy library
import time => current time uthata hai.

→ no need for memorizing the code => only understand concept

<u>Expected Outcome</u>

numpy se time kam aya, parallel se time zadha aya.

numpy ka algorithm boht effcient hota

~~Expo~~ GIL masla kita haii => multi threading mai time leta haii {process start, create, destroy mai bhi time lagta}

Numpy mai threading nahi karne hote

↳ Data bacha hona chahiye → for Best Result

MPP Massive Parallel Processing

Distributed hon or woh independent kaam kr rahe hon toh woh bhi MPP hai

use kete computers

Distributed memory => jab alg memory ko multiple

Shared memory => within computer

Difference important!

CRUD tool kit

conda create -n MyEnv

Task

text / image / ~. data~ csv form data

hospital / fashion data

environment create

do visualization → graphic banegi

reading time ? ; correlation find ?

kitne aise record hain ------>

→ GPU

google colab

select → Iris Dataset (download) → python mai load (read) → Classification/regression algorithm (SVM)

→ GPU

//

Book => Parallel and High Performance Computing

CLO's

- Basic concepts of parallel and distributed comput.
- algorithms
- GPU

Theory

Distributed System

A collection of independant computers that appear as a single system.
eg claud

Models of Distributed System

peer 2 peer — server
P2P => nodes directly communicate without a central

Enabling technologies for distributed system

protocol => expected outcome pata hotee, rules and regulations hotee
middle ware ⇐ ایک جگہ سے دوسری جگہ بات پہنچانا

Cloud and edge computing

Internet k through joh communication hotee processing se
Main Difference Edge mai processing zadha hai cloud mai time zadha lageh gaa kyu k woh doorahdoor servers pareh hain.
Edge mai hum qareb kre servers ko, distance less → access fast
Edge → synchronized hotee → automatic chra hotu remote or car ka server qareeb signals aram se or jaldi chale jate

Distributed file system => alag alag PC's ke fil ko manage kila.

Q) Cloud and Edge? Q) Scenario → what should be placed in above scenario?
Q) Diff b/w parallel and distributed with eg!

PDC

LAB → our concern is performance (how to improve) not to learn the code of classification

Commands:
- conda create --name env python=3.10
- conda activate env
- conda install -c conda

CPU or GPU ke performance dekhni
Library ⇒ Tensorflow
CPU or GPU ka aik dusre k sath supportive ho.

Procedure

1. Neural Network Training
   use (CIFAR-10) dataset
   training (feed kte baar baar) ⇒ use training loop
   testing
   optimization algorithm ✗
2. Comparision of CPU and GPU

Experiment 6 ہمارا کام ہے بس performance کا
- CPU per kitna time leta
- GPU
- synchronize CPU or GPU ko ke k time dekha?

Assignment
Collab file → CIFAR dataset
any algorithm → do classification
→ evaluation processing CPU time, GPU and synchronization ~~time~~ (hybrid)

THEORY

Why Parallel Computing?

جہاں challenge ہوتے وہاں پر کام کرتے

→ computational power jab zadha chahiye hotee

Parallel Computing
execution of many operations at a same time

→ Maximum computational power
→ Handle large problem
→ Simulation speed barha sakti

Serial computing => problem hotee hai or processor hota

Parallel computing => baree problem ko chote tukron mai torhna or phr solve kerna

ایسے tasks دیکھنے کے simultaneously کام کر سکتے
اگر ایک task دوسرے پر dependant ہے تو parallel computing نہیں ہو سکتی

Application and Results

less power consumption
· large data set per time zadha lagta toh energy zadha lagtee
energy efficiency کو نظر میں رکھتے ہوئے problem size اور system کی performance کو بڑھا کر تے ہیں

Why learn parallel computing?

· progress · performance
· serial performance archi nahi theee
· clock frequency barh gayi kyu k processors bhi barh gaye

Clock Cycle : instruction execute per second.
→ Computational power agar hum sahi se nahi leh rahe or hardware bht acha hai toh is ka koi faida nahi hai

Serial
do_payroll()
data aya → instruction mai split hua → processor
کے پاس گیا پھر

Parallel
tasks ko divide kiya

Amdahl's law ⇒ Definition imp!
Speedup (N) = 1 /(S + P/N)

fixed problem ⇒ depends ~~of~~ on no of processors

P ⇒ parallel, S ⇒ serial, N ⇒ no of processors

Limitation
serial ہم جتنی مرضی speed بڑھا لیں parallel کی مگر
part سے effect پڑتا ہے
performance per farq parta serial part se
~~Strong scaling~~ اگر serial اچھا ہو تو time کم لگے گا

Q) Definition? Limitations? Imp!!

~~Guast~~ Gustafson - Barss ⇒ N = N - S*(N-1)
problem size ↑ → processor ↑ بھی بڑھا دے → اس سے speedup بڑھ جائے گی

Benefits

size ↑ → divide → different processors
→ time ↑ => This improved scalability

weak scaling => time zadha lagta
execution of proces mai joh time lagta
w.r.t no of processors mai lagta.

Strong :- problem divided → fixed across processor
→ speed up calculation.

Parallel computing Slide 13

layers

- source code
- compiler
- OS

manages execution on hardware.

Data vectorization

aisa koi fnaam hein joh sab data
per apply hagaye
eg matrix => 3 x 3 => overall matrix # 1 krdo

Benefits

① faster computation
② optimized for array operation (eg NumPy)

P5

PDC

Categorizing Parallel approaches

kese categorize kr sakte.

ایک time پر ایک ہی data execute ہو۔

SIMD => GPU mai hotee zada

MISD => space craft mai hota.

fault-tolerant => کوئی fault آتا تو OS اس کو کیسے cover کر لیتا

MIMD => data or instruction ke parallelism

SIMT => different data pe operations hote.

~~Parallel speedup vs Comparative Speedup~~

Content

Problem ka pata hona chaiye, toh usi se hum achi performance kr sakte.

Assignment => stream processing
4 approaches => detail
کب کون سی use کرنی ہے