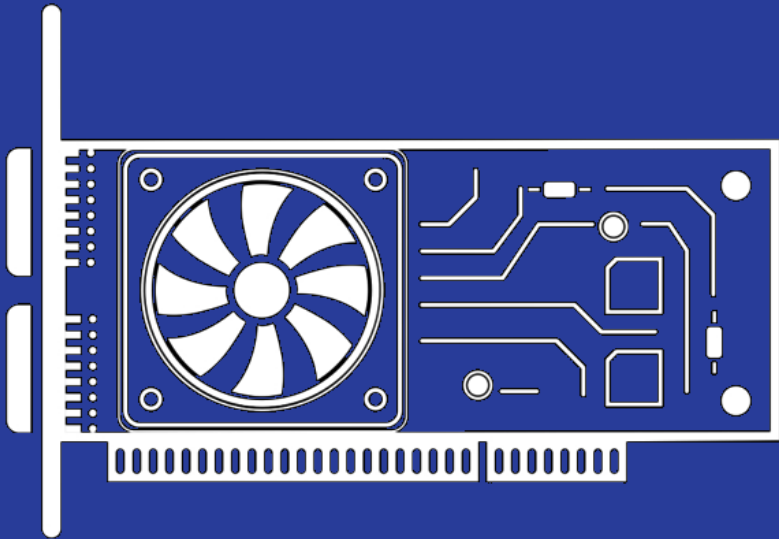# Advanced Computer Architectures

Nitin Tiwari

- Alumni | 2017-2021 | Information Technology
- Software Engineer at LTI
- Blog Publisher, TFUG Mumbai
- Contribute to the TensorFlow Community

April 06, 2022

Contents

- History of Computer Generations

- Advanced Computer Architecture

- Capabilities of Advanced Computer Architecture

- Graphical Processing Unit (GPU) & Tensor Processing Unit (TPU)

- Q&A (feel free to shoot in your queries)

History of Computer Generations

Zeroth Generation (1642)
Pascal's Calculator
Performs simple calculations
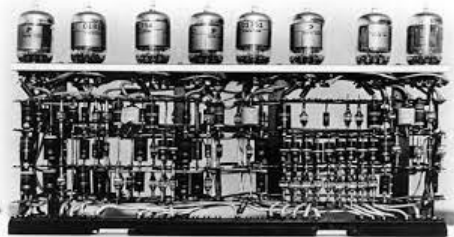
First Generation (1940-1950s)
Vacuum Tube-based
Machine Language, large in size
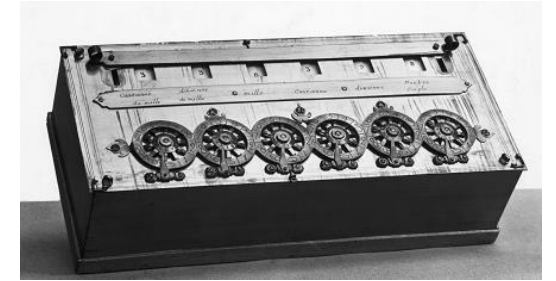
Second Generation (1950-1960s)
Transistor-based
Assembly language

Third Generation (1960-1970s)
IC-based
Mini-computers

Fourth Generation (1970 - present)
Microprocessor-based
VLSI – 1000s of transistors/single chip
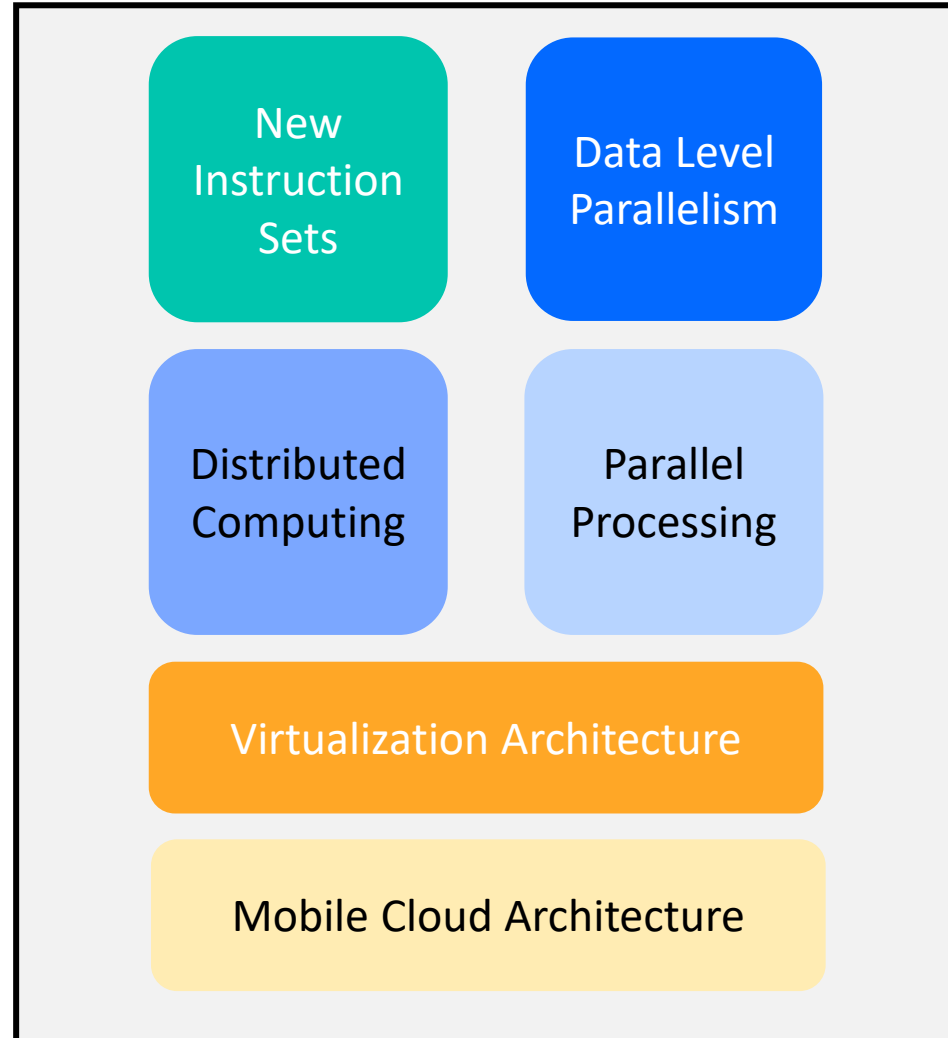
Fifth Generation (Present and future)
Artificial Intelligence-based
ULSI – millions of transistors/chip

Advanced Computer Architectures

New Instruction Sets

Data Level Parallelism

Distributed Computing

Parallel Processing

Virtualization Architecture

Mobile Cloud Architecture

Advanced Computer Architectures

New Instruction Set:

Complex programs can be written using a new set of instructions that has the capability to handle the logics that require more time and computation otherwise.

Example:
Write a program to swap two variables.

a = 10, b = 20

Traditional approach:
Step 1: Define a new variable c.
Step 2: Assign c := a
Step 3: Assign a := b
Step 4: Assign b := c

Advanced approach:
Step 1: a, b = b, a
That's it.

Data Level Parallelism

When a single operation has to be performed on multiple data elements, data-level parallelism can be used to reduce computation time and effort.
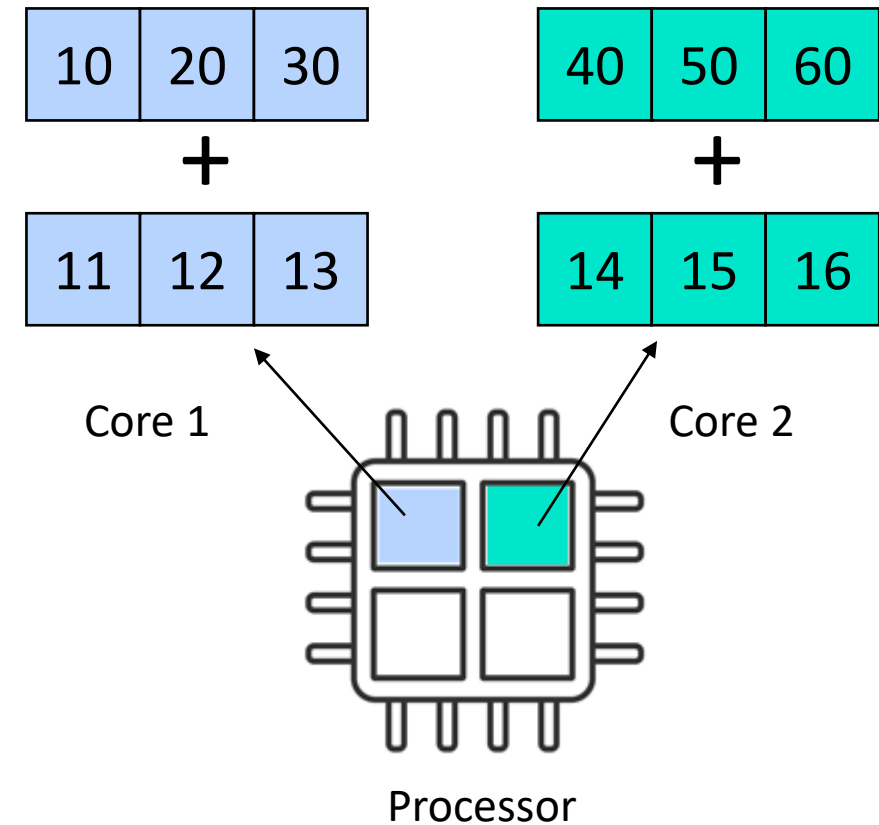
Example: Calculate C[i] = A[i] + B[i]

A =

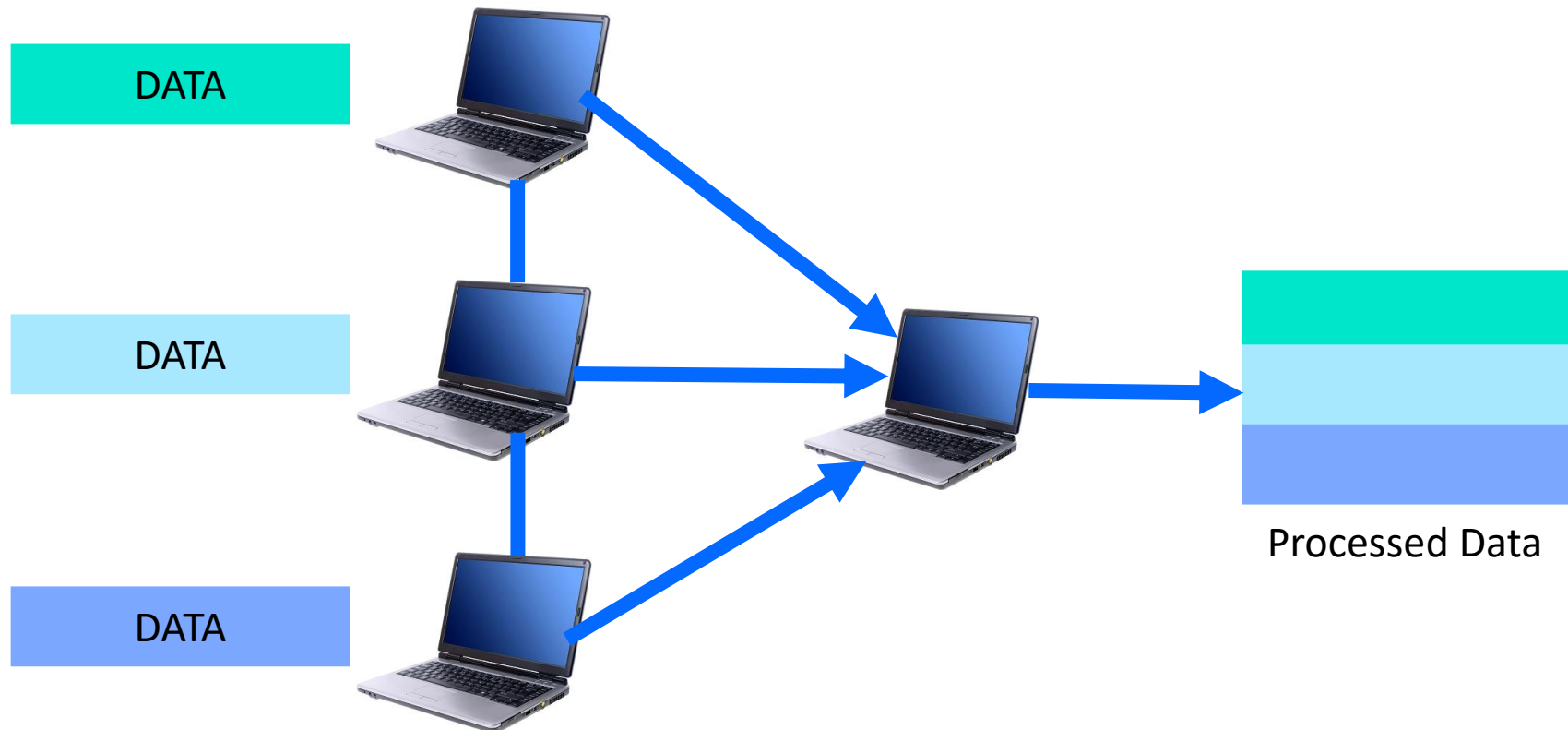| 10	20	30	40	50	60

B =

11	12	13	14	15	16

10	20	30

+

11	12	13

Core 1

40	50	60

+

14	15	16

Core 2

Processor

Distributed Computation

- Distributed Computation lets you compute/process your data on different computers over a network into a cluster.

- The processed data from these different systems are then combined together.

Processed Data

Parallel Computing

A real-world example of parallel computing using AI:

Processes happening in the above application parallelly:

- P1: Fetching frames from the webcam.
- P2: Processing the frames through a Neural Network to detect and classify poses.
- P3: The computer hits the key on the keyboard.
- P4: The player plays a shot in the game.

Virtualization

- Modern computer architectures enable the feature of virtualization of resources.
- Resources can be anything, including operating system, server, network, storage, etc.

For example:
- Android Studio, a software used for developing Android apps, supports Android Virtual Devices (AVDs) which can be run on Windows/Linux operating system.
- VirutalBox allows you to run one OS on top of another OS.
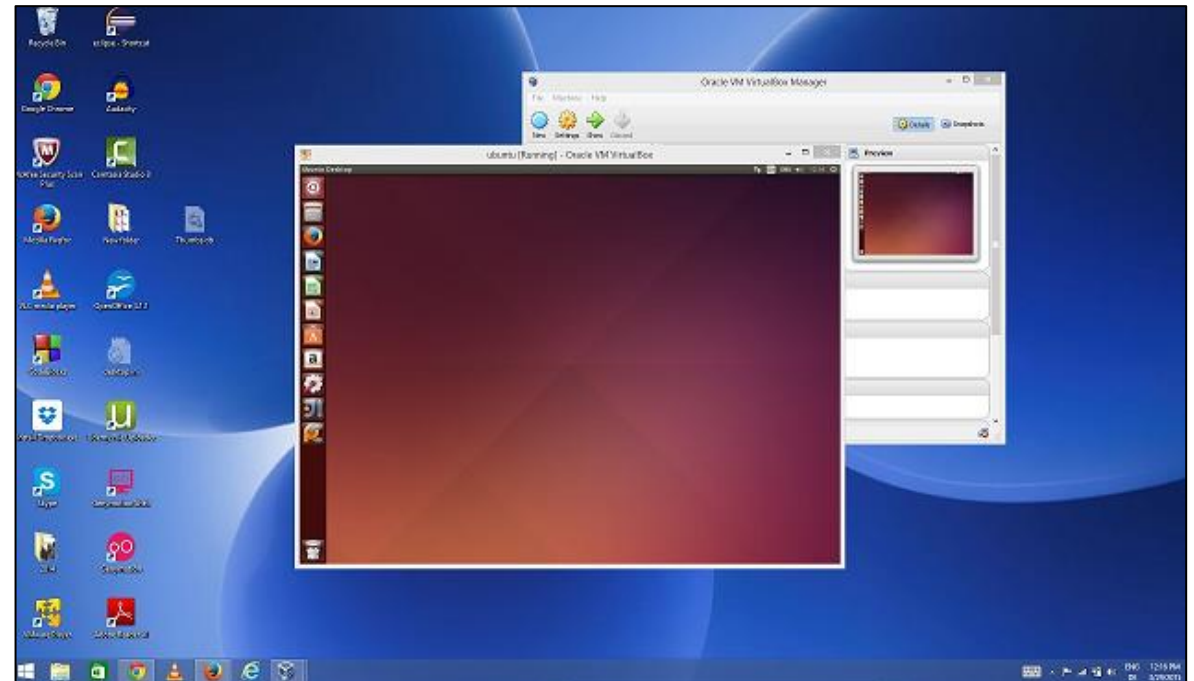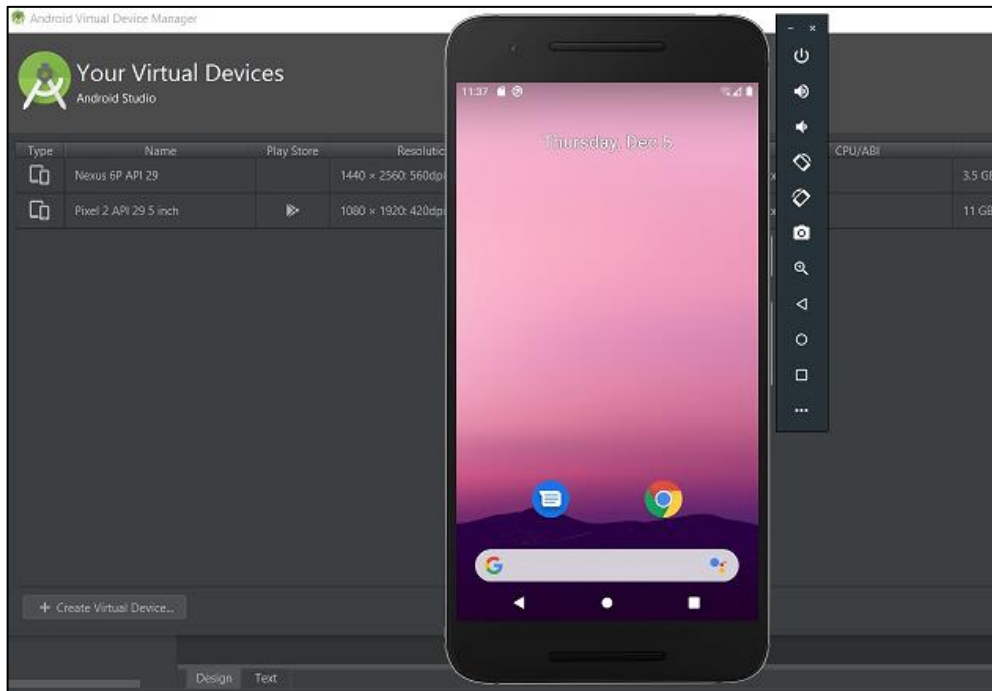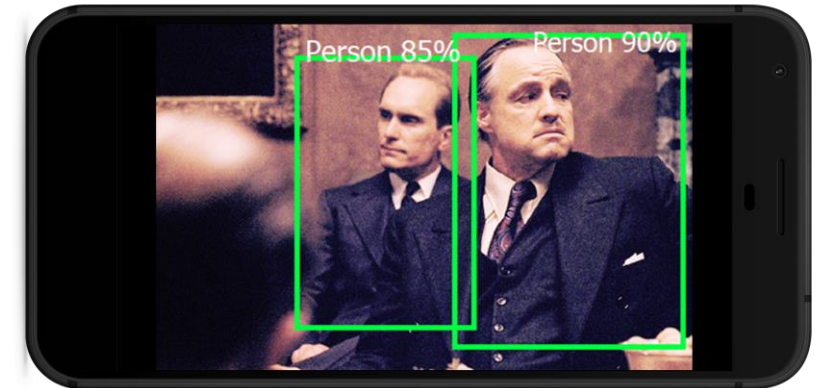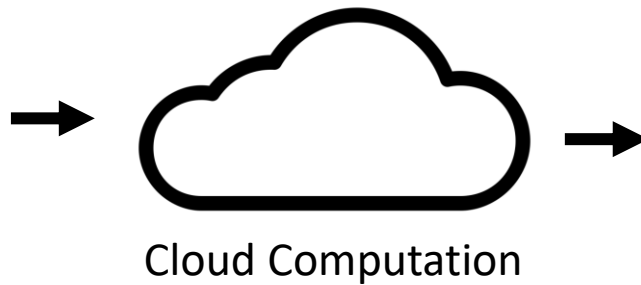
Mobile Cloud Computing

- Mobile Cloud Computing has been revolutionary in the uprise of modern computer architectures.

- In MCC, computations, and processing of the data happen on the cloud and the results are displayed on mobile devices for the users to use.

Example:

Object detection on mobile applications.

Data   Cloud Computation   Object Detection results on mobile device

If you remember this game, your childhood was awesome.

Super Mario Bros – 32 KB

(Fun fact: This presentation is larger in size than this game)

Capabilities of Advanced Computer Architecture

Call of Duty: Modern Warfare (2019) – 175 GB

- Advanced Computer Architecture has made it possible to run high-end graphics games.

- Hundreds and thousands of processes run parallelly to process the data (frames).

TensorFlow

- Free, open-source Machine Learning library.

- Originally developed by the Google Brain team.

- Languages supported: Python, C++, Java, JavaScript.

- Official GitHub repository: www.github.com/tensorflow

- Official website: www.tensorflow.org

Scalar: 0D data

1

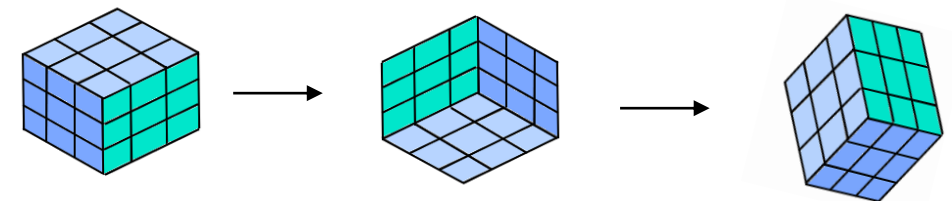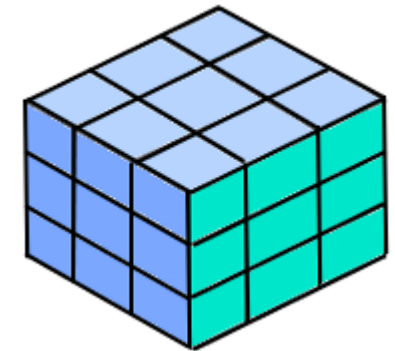**Vector:** 1D data

| 1 | 2 |

Matrix: 2D data

| 1 | 2 |
| 3 | 4 |
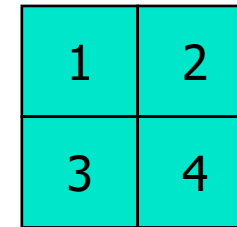
Tensor: N-D data

The flow of tensor data in a neural network.
Hence, the name – 'TensorFlow'.

GPU & TPU

- Graphical Processing Units (GPUs), are specialized processors that are designated to accelerate graphics rendering.

- As the name suggests, it is mostly used for graphical applications.

- They can process a lot of data simultaneously and are thus used in Machine Learning, video editing, and gaming applications.

- Tensor Processing Units (TPUs) are customized processors developed by Google for processing large tensor data for ML applications.

Central Processing Unit (CPU)

Graphical Processing Unit (GPU)
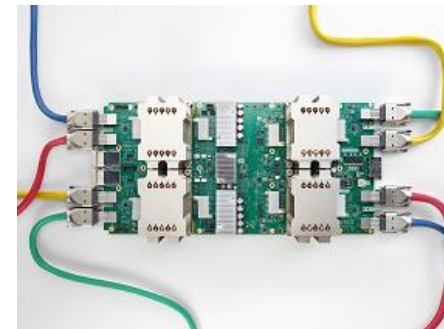
Tensor Processing Unit (TPU)

www.linkedin.com/in/tiwari-nitin

tiwarinitin1999@gmail.com

www.tiwarinitin1999.medium.com

www.github.com/NSTiwari