Ch-2

Performance Issues

i) What are the different methods used to increase the microprocessor speed?

ii) What do you mean by pipelining, its advantages

iii) Why do you need to achieve the performance balance?

iv) What are the diffⁿ obstacles we face, while increasing the processor speed?

v) Multicore processors, MIC & GPGPUs

vi) How the clock pulses are generated?

vii) Define ~~Divide~~ the terms clock rate, clock cycle & cycle time.

viii) Problems from CPI, MIPS Rate, MFLOPS Rate, T, AM, GM, HM

Normalized value, speed of a processor, Define benchmark

Microprocessor Speed

With the following methods, we can increase the microprocessor speed

1) **Pipelining:-**

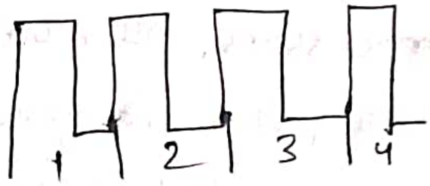
- Pipelining enables the processor to work simultaneously on multiple operations at the same time by performing the different phases of an instruction execution cycle, i.e. the phase cycle & the execute cycle.
- Pipelinging improves the efficiency of the processor

(i) ADD 0123H, 1234H

Opcode Fetch Cycle — Execute Cycle (T)

(ii) SUB

(iii) MOV



With pipelining	Without pipelining
1cc → F1	1cc → F1
2cc → E1+F2	2cc → E1
3cc → E2+F3	3cc → F2
4cc → E3	4cc → E2
	5cc → F3
	6cc → E3

2) **Branch Prediction :-**

The processor looks ahead in the instruction code faced from memory & predicts branches or groups of instructions are likely to be processed next

3) **Data Flow Analysis :-**

The processor analyzes which instructions are dependents on each others result on data to create and optimised schedule instruction.

4> Superscalar Execution:-
This is the ability of the processor to execute more than one instruction in a single clock cycle, so that multiple parallel pipelines can be executed.

5> Speculative Analysis:-
~~The~~ With the help of Branch prediction & Data flow analysis some of the processor can theoretically execute information prior to their actual appearance. (This is the basis of AI & ML)

6> By increasing the RAM Size:-

7> By increasing the cache memory

8> By increasing the chip density

9> By increasing Multi-core processors.

Performance Balance

- Performance balance refers to the adjustment of the organization & architecture to compensate for the speed mismatch among the different structural component.
- The main performing ~~some of the performance~~ affecting parameters are:-
  (i) Interface b/w the processor and main memory
  (ii) ~~Handling~~ Handling the I/O devices
  (iii) Balance the throughput (efficiency) and the processing demand of the processor components & also the interconnection structures

Dt-29-02-24

- Obstacles that we face while improving the performance of the processor

1> Power
- As the density of the chip increases & also the clock speed increases, so also the power density increases. Therefore we require good heat decepating electronic circuits which is very difficult to fabricate on the high density chips.

Dt-29-02-24

2) RC Delay :-

delay $= \tau = 1/RC$

Delay increases as RC product increases

3) Memory Latency and throughput :-

Memory access speed (latency) & transfer speed (throughput) lags as the processor speed increases

* Multicore, MIC & GPGPUs :-

MIC :- Many Integrated Core

MIC architecture is a term used by Intel which refers to a series of microarchitectures that integrated many physical cores onto a single integrated circuit.

GPGPUs :- General Purpose Graphic Processing Unit

GPU are specialised cores used for video processing

Ex :- All PCs, PlayStation, Mobile Phones

* Two Laws :-

[They provide insight into the performance of processors]

1. Amdhal's Law
2. Little's Law

1. Amdhal's Law

- Speed up in one aspect doesn't result in a corresponding improvement in the performance (increasing the no. of cores) & this limitation is expressed by Amdhal's Law
- Consider a program running on a single processor such that a fraction $(1-f)$ of the execution time involves codes that are inherently sequential in nature and a fraction f that is infinitly parallel in nature. So let T be the total execution time of the program using a single processor then the

speed up by using parallel processor with N no. of processors that completely executes the parallel portion of the program is given as

$$\text{speed up} = \frac{\text{Time to execute a program in single processor}}{\text{Time to execute a program on n parallel processor}}$$

$$\boxed{\text{Speed up} = \frac{1}{(1-f) + f/N}}$$

- Two important conclusions of Amdhal's Law
  (i) If f is small, the use of parallel processors has little or no effect on the speed up.
  (ii) ~~If f~~ As N approaches ∞, speed up will be only depending on

$$\frac{1}{1-f}$$

Dt - 02 - 03 - 23

$$\text{Speed up} = \frac{\text{Performance after enhancement}}{\text{Performance before enhancement}}$$

$$\text{Speed up} = \frac{\text{Execution time before enhancement}}{\text{Execution time after enhancement}}$$

$$\text{Speed up} = \frac{1}{(1-f)+(f/su_f)}$$

Here f = fraction of time before enhancement
su_f = speed up of the feature after enhancement

Q) Suppose a task makes extensive use of floating point operation with 40% of the time consumed by floating point operation. In a new hardware design the floating point module is sped up by factor k. Compute the overall speed up.

Ans:- Given.
$f = 40\% = 0.4$

$su_f = K$

We can solve the speed up by ~~using~~ assuming

K = 100

$$\text{Speed up} = \frac{1}{(1-0.4)+(0.4/100)} = \frac{1}{0.6+\left(\frac{0.4}{100}\right)} = 1.6556$$

K = 200

$$\frac{1}{(1-0.4)+200} = \frac{1}{200.06}$$

K = 200

$$\text{Speed up} = \frac{1}{(1-0.4)+(0.4/200)} = \frac{1}{0.6+0.002} = \frac{1}{0.602} = \cancel{1.658}\ 1.66$$

K = 400

$$\text{Speed up} = \frac{1}{(1-0.4)+(0.4/400)} = \frac{1}{0.6+0.001} = \frac{1}{0.601} = 1.663$$

Speed is independent of the speed up factor values

Q) The execution time of half of the program can be accelerated by factor 2. What is the overall program speed up.

Given, $f = \frac{1}{2} = 0.5$

$K = 2$

$$\text{Speed up} = \frac{1}{(1-0.5)+\left(\frac{0.5}{2}\right)} = \frac{1}{0.5+0.25} = \frac{1}{0.75} = 1.33$$

Q) $f = 0.1$

& $K = 10$

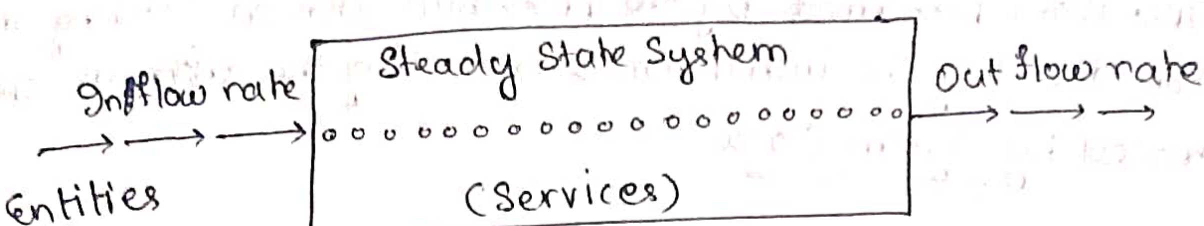
$$\text{Speed up} = \frac{1}{0.9+\frac{0.1}{10}} = \frac{1}{0.91} = 1.098 \approx 1.1$$

Q) $f = 0.9$

$K = 10$

$$\text{Speed up} = \frac{1}{0.1+0.09} = \frac{1}{0.19} = 5.26$$

2> Little's Law :-



We have a Steady State System, where items arrive at an average rate of 'λ' items per unit time. The items stay in the system for an average of W units of time. Finally there is an average of L units in the system at one time.

- Little's Law relates all these variables $\boxed{L = \lambda W}$.

Q> A doctor in the hospital observes that on an average 6 patients per hour arrive and there are typically 3 patients in the hospital. What is the average length of time each patient spends in the hospital.

Ans-

Given, $L = 3$

$\lambda = 6$

$W = L/\lambda = 3/6 = 0.5$ hr. (Ans)

Q> An out patient clinic works 12 hrs a day. On an avg 542 patients visit the clinic each day and ~~seek~~ seek consultation & each patient spends 28 minutes in the clinic. How many patients are present in the clinic on an avg at any given time during working hours?

$L = \lambda * W = \left(\frac{542}{12*60}\right)$ per minutes $\times 28$

$\lambda = 542$ over 12 hrs

$W = 28$ minutes (avg time spent)

$L = 21$ patients (Ans)

Q> A bread manufacturing line typically produces 150 loaves of bread during 8 hrs shift. On average, there are 30 loaves (equivalent) under processing in the line at any given time. What is the average time it takes to produce a loaf of bread from the raw ingredients.

$W = L/\lambda$

$= 30/(150/8)$ hrs

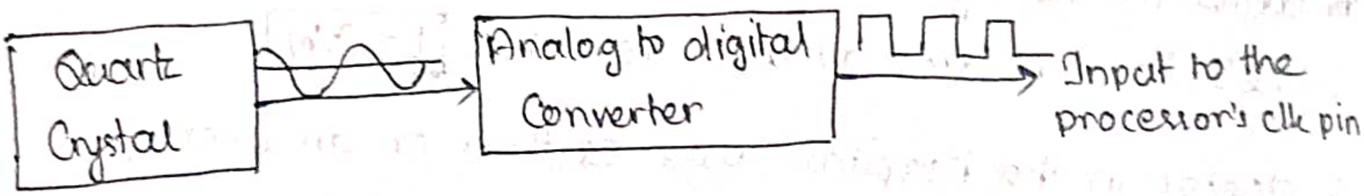
$= 1.6$ hrs (Ans)

* Basic Measures of Computer Performance :-

1) Clock Speed :-

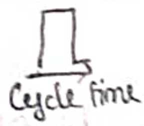
- Operations performed by the processor such as fetching an instruction, decoding the instruction & executing the arithmatic operatio is controlled by System Clock.

- Generally all the operation start with the clock ~~calls~~ pulse.
- The speed of the processor is decided by the clock pulse frequen & is measured in Hz/second
- Method of generating clock pulse



(System Clock) :

- Clock signals are generated by a Quartz Crystal, which produce const sign waves. This wave forms are converted into digital signa or clock pulses, with the help of a ADC.
- The rate of pulses is known as clock rate / clock speed.
- One increment or clock pulse is known as Clock cycle / Clock tick
- The time b/w the pulses is known as Cycle time.



2) Instruction Execution Rate :-

(i) Frequency
→ A processor is ~~const~~ driven by of clock with a const frequency(f) or equivalently a constant cycle time (τ) such that

$$\boxed{\tau = 1/f}$$

Q) Find out the value of cycle time, if a processor frequency is 400MHz

Given, $f = 400 MHz = 400 \times 10^6 Hz$

$\tau = 1/400 \times 10^6 = 2500$ sec

DT - 05-03-24

(ii) Instruction Count (I_c)

Instruction Count for a program is defined as the no. of processor instruction executed for that program till its complition for a defined time interval.

(iii) CPI (Cycles Per Instruction) (Avg. Cycles per Instruction for a program)

Why the CPI is an average quantity?

Ans- CPI is an average quantity because it represents the average no. of clock cycles required to execute one instruction across a diverse set of instructions within a program.

- Let CPI_i be the no. of cycles required for instruction type i & I_i be the no. of executed information for a given program. Then the overall CPI is given as

$$\boxed{CPI = \frac{\sum_{i=1}^{n} (CPI \times I_i)}{I_c}}$$

(iv) T - The processor time T needed to execute a given program can be expressed as

$$\boxed{T = I_c \times CPI \times \tau}$$

(v) MIPS rate (Millions of Instruction per second)

It is a common nature of performance for a processor & it defines the rate at which the instructions are executed.

$$\boxed{\text{MIPS Rate} = \frac{I_c}{T \times 10^6} = \frac{f}{CPI \times 10^6}}$$

f) MFLOPS rate (Millions of Floating point operations Per Second)

It deals with floating point instruction. They are normally used in scientific & gaming application.

$$\text{MFLOPS rate} = \frac{\text{No. of executed floating point operation in a program}}{\text{Execution time} \times 10^6}$$

Q What is the CPI of a program execution that consists of the following instruction classes?

Instruction Class	CPI	% mix (I_i)
A	1.4	25
B	2.4	70
C	2	05

$$CPI = \sum_{i=1}^{n} \frac{CPI_i \times I_i}{I_c} = \frac{1.4 \times \frac{25}{100} + 2.4 \times \frac{70}{100} + 2 \times \frac{5}{100}}{100/100} = 2.13$$

Q Consider the execution of a program that results in the execution of 2 million instructions on a 400 MHz processor. The program consists of 4 main types of instructions. The instruction ~~makes~~ mix and CPI for each instruction type is given below based on the result of a program trace experiment. Calculate the average CPI & the corresponding MIPS rate,

Instruction Type	CPI	Instruction mix (I_i)
Arithmatic · logic	1	60
Load/store with cache hit	2	18
Branch	4	12
Memory reference with cache miss	8	10

$$CPI = \sum_{i=1}^{n} \frac{CPI_i \times I_i}{I_c} = \frac{1\times\frac{60}{100} + 2\times\frac{18}{100} + 4\times\frac{12}{100} + 8\times\frac{10}{100}}{100/100}$$

$$= \frac{0.6 + 0.36 + 0.48 + 0.8}{1} = 2.24$$

$$\text{MIPS Rate} = \frac{100/100}{T\times 10^6} \quad \frac{f}{CPI\times 10^6} = \frac{400\times 10^3}{2.24\times 10^6} = 178.57\times 10^{-3}$$

~~$T = I_c \times CPI \times Z = 1\times 2.24\times \frac{1}{\ }$~~

Q Suppose one processor A executes a program with an average CPI of 1.9. Suppose another processor with the same instruction set and an enhance compiler, execute the same program with 20% less instruction & with a CPI of 1.1 at 800MHz. In order for the 2 processors to have the same performance, what does the clock rate of the 1st processor need to be?

Ans -

A

$CPI = 1.9$

$f_A = ?$

B

20% less ⇒ 0.2 less

$CPI = 1.1$

$f = 800\,MHz$

~~$MIPS = \frac{800}{1.1\times 10^6}$~~

$T = I_c \times CPI \times Z$

$= I_c \times CPI \times \frac{1}{f}$

$\therefore T_A = T_B$

$\Rightarrow I_c \times CPI \times \frac{1}{f_A} = I_c \times CPI \times \frac{1}{f_B}$

$\Rightarrow x \times 1.9 \times \frac{1}{f_A} = x(1-0.2)\times 1.1 \times \frac{1}{800\times 10^{-6}}$

$\Rightarrow \frac{1.9}{f_A} = 0.8 \times 1.1 \times 1.25 \times 10^{-9}$

$\Rightarrow f_A = \frac{1.9}{0.8\times 1.1\times 1.25\times 10^{-9}}$

$= \frac{1.9}{1.1\times 10^{-9}}$

$\Rightarrow f_A \simeq 1.72\,GHz$

Q) Which is the best way (Clock rate, CPI, MIPS rate, MFLOPS Rate, Memory Latency or average execution time) to measure the performance of a processor? & why?

Ans- The best way to measure the performance of a processor depends on the specific goals of the evaluation.

Ex:-

i) For architectural efficiency and instruction execution, CPI may be the most relevant metric.

ii) For evaluating memory subsystem performance, memory latency is crucial.

iii) Real world performance analysis may prioritize metrices like average execution time, which considers the overall time taken to complete tasks.

Q) What is the average CPI of the 1.4 GHz processor that executes 12.5 millions instructions in 12 seconds.

Ans- Given, $f = 1.4\,GHz = 1.4 \times 10^{12}$

$$MIPS = \frac{12.5}{12} = 1.04$$

$$MIPS\ rate = \frac{f}{CPI \times 10^6}$$

$$1.04 = \frac{1.4 \times 10^{12}}{CPI \times 10^6}$$

$$\Rightarrow CPI = \frac{1.4 \times 10^{12}}{1.04 \times 10^6} = 1.34 \times 10^6$$

$$\therefore \boxed{CPI = 1.34 \times 10^6}$$

Calculating the mean :-

Mean gives us the speed of the processor

(i) (AM) ~~Arithmal~~ Arithmatic mean :- $\dfrac{x_1 + x_2 + x_3 + \dots + x_n}{n}$

(ii) (GM) Geometric Mean :- $\sqrt[n]{x_1 . x_2 . x_3 \dots x_n}$

(iii) (HM) Harmonic Mean :- $\dfrac{n}{\left(\frac{1}{x_1}\right) + \left(\frac{1}{x_2}\right) + \dots + \left(\frac{1}{x_n}\right)}$

$$\boxed{AM > GM > HM}$$

This equality is always true

DT - 07-03-24

Q. In the table given below a comparison b/w the performance of 3 computers, i.e. computer A, B, C on execution of 2 programs

	Comp A (in sec)	Comp B (in sec)	Comp C (in sec)
Program 1 (10^8 FP oper)	2.0	1.0	0.75
Program 2 (10^8 FP opr)	0.75	2.0	4.0
Total execution ~~Program 3~~ time	2.75	3.0	4.75

Calculate a) ~~10^8 FP operation~~ the arithmatic mean of time for each computer

b) inverse of total execution time.

c) MFLOPS rate for each computer, for each program taking into account that the execution of each program results in the execution of 10^8 FP operations.

d) The arithmatic mean of MFLOPS rate per each computer

e) Harmonic mean of rates for each computers.

f) Write the conclusion.

Ans

a) Comp A = $\frac{2.75}{2} = 1.375$

Comp B = $\frac{3.0}{2} = 1.5$

Comp C = $\frac{4.75}{2} = 2.375$

b) Comp A = $\frac{1}{2.75} = 0.363\ sec^{-1}$

Comp B = $\frac{1}{3.0} \doteq 0.33\ sec^{-1}$

Comp C = $\frac{1}{4.75} = 0.210\ sec^{-1}$

c) Prog 1 ~~Comp A~~ MFLOPS Rate

Comp A ~~Prog 1~~ = $\frac{10^8}{2 \times 10^6} = 0.5 \times 10^2 = 50$

Comp B ~~Prog 2~~ = $\frac{10^8}{1 \times 10^6} = 100$

Comp C ~~Prog 3~~ = $\frac{10^8}{0.75 \times 10^6} = 133.3$

Prog2 MFLOPS

Comp A = $\frac{10^8}{0.75 \times 10^6} = 133.3$

Comp B = $\frac{10^8}{2 \times 10^6} = 50$

Comp C = $\frac{10^8}{4 \times 10^6} = 25$

d) Comp A = $\frac{50 + 133.3}{2} = \frac{183.3}{2} = 91.65$

Comp B = $\frac{100 + 50}{2} = 75$

Comp C = $\frac{133.3 + 25}{2} = 79.15$

e) HM of MFLOPS

Comp A = $\frac{2}{\left(\frac{1}{50}\right) + \frac{1}{(133.3)}} = \frac{2}{0.02 + 0.0075} = \frac{2}{0.0275} = 72.72$

Comp B = $\frac{2}{\left(\frac{1}{100}\right) + \left(\frac{1}{50}\right)} = \frac{2}{0.01 + 0.02} = \frac{2}{0.03} = 66.66$

Comp C = $\frac{2}{\left(\frac{1}{133.3}\right) + \left(\frac{1}{25}\right)} = \frac{2}{0.0075 + 0.04} = \frac{2}{7.54} = 42.11$

7.54×10^{-1}

f) Conclusion Table

	Comp A (in sec)	Comp B (in sec)	Comp C (in sec)	MFLOPS Rate (A)	MFLOPS Rate (B)	MFLOPS Rate (C)
Program 1 (10^8 FP operation)	2.0	1.0	0.75	80	100	133.3
Program 2 (10^8 FP operation)	0.25	2.0	4.0	133.3	50	25
Total execution time	2.75	3.0 A>B>C	4.25	–	–	–
a) AM of times	1.33	1.5 A>B>C	2.32	–	–	–
b) Inverse of Exe time	0.36	0.33	0.21	–	–	–
c) AM of rate	–	–	–	91.67	75.00 A>C>B	79.15
d) HM of rates	–	–	–	72.72	66.66 A>B>C	42.11

- From the total execution time we conclude that A is faster than B than C (A > B > C)
- From the AM values of times we can also conclude the ~~same~~ A > B > C
- From the AM of rates we are getting A > C > B which is a mismatch

* HM always gives more accurate result than AM

* HM is inversely proportional to the total execution time - desired property

Q)

	Comp A (in sec)	Comp B (in sec)	Comp C (in sec)
Program 1 (10^8 FP opr^n)	2.0	1.0	0.75
Program 2 (10^8 FP opr^n)	0.75	2.0	4.0
Total Execution Time	2.75	3.0	4.75

$\boxed{\text{GM exhibit consistency for Normalized value}}$

Calculate a) all normalized values with respect to comp A (AM, GM)
b) all normalized values w.r.t comp B (AM, GM)

a) Normalized w.r.t Comp	Comp A	Comp B	Comp C
Program 1 (10^8 FP operation)	$\left(\frac{2.0}{2.0}\right) = 1$	$\left(\frac{1.0}{2.0}\right) = 0.5$	$\left(\frac{0.75}{2.0}\right) = 0.375$
Program 2 (10^8 FP opr^n)	$\left(\frac{0.75}{0.75}\right) = 1$	$\left(\frac{2.0}{0.75}\right) = 2.66$	$\left(\frac{4.0}{0.75}\right) = 5.33$
Total Execution Time	2.75	3.0	4.75
AM of normalized values	1	1.5	2.85
GM of normalized values	1	1.15	1.41

- From the total execution time, we can conclude that A>B>C
- From the AM of normalized value A<B<C

Normalized value	Comp A	Comp B	Comp C
Program 1 (10^8 FP oprⁿ)	$\frac{2.0}{1.0} = 2$	$\frac{1.0}{1.0} = 1$	$\frac{0.75}{\cancel{0.5} \; 1.0} = \cancel{1.5} \; 0.75$
Program 2 (10^8 FP oprⁿ)	$\frac{0.75}{2.0} = 0.375$	$\frac{2.0}{2.0} = 1$	$\frac{4.0}{2.0} = 2$
Total execution time	2.75	3.0	4.75
AM of normalized values	1.187	1	1.375
GM of normalized values	0.866	1	1.22

Dt - 09 - 03 - 24

Q

	Comp A	Comp B	Comp C
Prog 1	2.0	1.0	0.20
Prog 2	0.4	2.0	4.0

For the table given above calculate the normalized values w.r.t (a) Comp A (AM, GM)

(b) Comp B (" , GM)

(c) Comp C (AM, GM)

Normalized Value w.r.t Comp A	Comp A	Comp B	Comp C
Prog 1	1	0.5	0.1
Prog 2	1	5	10
Total Execution Time	2.4	3.0	4.2
AM of normalized Val	1	2.75	5.05
GM of normalized Value	1	1.11	3.16

Normalized value w.r.t Comp B	Comp A	Comp B	Comp C
Program 1	2	1	0.2
Program 2	0.2	1	2.0
Total Execution Time	2.4	3.0	4.2
AM of normalized value	1.1	1	1.1
GM of normalized value	0.28	1	0.28

Normalized value w.r.t Comp C	Comp A	Comp B	Comp C
Program 1	10	5	1
Program 2	0.1	0.5	1
Total Execution Time	2.4	3.0	4.2
AM of normalized value	5.05	2.52	1
GM of normalized value	0.31	1.11	1

Q A Benchmark program is run ~~1st~~ fast on a 200MHz processor & a 300MHz processor. Executed program consists of 1M execution with the following instruction makes & a clock cycle count. Calculate the

(a) Effective CPI

(b) MIPS Rate for both the processor

Instruction Type	Instruction count	Cycles / Instruction
Integer Application	4,00,000	1
Data Transfer	3,50,000	2
Floating Point	2,00,000	3
Control Transfer	50,000	2

Ans - ~~Integer Application~~ effective CPI = (1×4,00,000)+(2×3,00,000)

a) $\text{Effective CPI} = \dfrac{1\times10^{6} + 2\times10^{6} + 3\times10^{6} + 2\times10^{6}}{4,00000 + 3,50000 + 2,00,000 + 50,000} = 1.8$

b) $\text{MIPS Rate of 1st program} = \dfrac{200\times10^{6}}{1.8\times10^{6}} = 111.11$

$\text{MIPS Rate of 2nd Program} = \dfrac{300\times10^{6}}{1.8\times10^{6}} = 166.66$

BenchMark :-

- Benchmarks are programs that are primarily used to compare one processor with another processor in terms of performance values.
- GM is always used for benchmark analysis as it provides consistent result, when measuring the relative performance of processors.

Advantages of benchmark:-

- Performance Enhancement
- Competitive Edge
- Fostering innovation & learning
- Informed strategic decision-making
- Customer Focus
- Collaboration & Networking
- Driving Organisational Change
- Effective Performance Measurement.