U23AIT33 - COMPUTER ORGANIZATION AND ARCHITECTURE

PART – A QUESTION BANK

UNIT I - OVERVIEW & INSTRUCTIONS

1.Define Computer.

A computer can be defined as a fast electronic calculating machine that can accept digitized data as input, process the data and produce information as output.

2. List down the Eight Great Ideas in Computer Architecture.

1. Design for Moore‘s Law
2. Use Abstraction to Simplify Design
3. Make the Common Case Fast
4. Performance via Parallelism
5. Performance via Pipelining
6. Performance via Prediction
7. Hierarchy of Memories
8. Dependability via Redundancy

3. How will you improve performance via parallelism?

Computer architects have offered designs that get more performance by performing operations in parallel. Use multiple jet engines of a plane for **parallel performance**.

4. what is the need for Hierarchy of Memories?

Programmers want memory to be fast, large, and cheap. The hierarchical Arrangement of storage in current computer architectures is called the memory hierarchy. It is designed to take advantage of memory locality in computer programs. Each level of the hierarchy is of higher speed and lower latency, and is of smaller size, than lower levels.

5. List the Components of a Computer System.

The underlying hardware in any computer performs the same basic functions: inputting data, outputting data, processing data, and storing data. The five classic components of a computer are

1. Input
2. Output
3. Memory

4. data path and control (with the last two sometimes combined and called the processor)

6. What you meant by Operating System?

An operating system interfaces between a user‘s program and the hardware and provides a variety of services and supervisory functions.

7. Define Moore’s Law.

Moore's law is the observation that, over the history of computing hardware, the number of transistors on integrated circuits doubles approximately every 18 to 24 months. The law is named after Intel co-founder Gordon E. Moore, who described the trend in his 1965 paper.

8. What you meant by Redundancy?

Redundancy is the duplication of critical components or functions of a system with the intention of increasing reliability of the system.

9. Define pixel.

The smallest individual picture element. Screens are composed of hundreds of thousands to millions of pixels, organized in a matrix.

10. What is the use of data path and control?

The processor logically comprises two main components: data path and control, the respective brawn and brain of the processor. The **data path** performs the Arithmetic operations, and **control** tells the data path, memory, and I/O devices what to do according to the wishes of the instructions of the program.

11. What are the main functions of memory unit and its types?

The functions of memory unit are to store programs and data. There are two classes of storage, they are Primary storage and Secondary storage.

12. Define RAM.

Memory in which any location can be reached in a short and fixed amount of time after specifying its address is called random-access memory (RAM).

13.Define addressing modes and its various types. (nov/dec 2017)

The different ways in which the location of a operand is specified in an instruction is referred to as addressing modes. The various types are Immediate Addressing, Register Addressing, Based or Displacement Addressing, PCRelative Addressing, Pseudodirect Addressing.

14.Write Basic performance equation in terms of instruction count (the number of instructions executed by the program), CPI, and clock cycle time.

$$\text{CPU time} = \text{Instruction count} \times \text{CPI} \times \text{Clock cycle time}$$

or, the clock rate is the inverse of clock cycle time:

$$\text{CPU time} = \frac{\text{Instruction count} \times \text{CPI}}{\text{Clock rate}}$$

15. What is system software and their functions?

System software is a collection of programs that are executed as needed to perform functions such as Receiving and interpreting using commands. Entering and editing application programs and storing them as files in secondary storage devices.

16.What are the types of instruction in MIPS. (APR/MAY2018)

1. Arithmetic instruction
2. Data transfer Instruction
3. Logical Instruction
4. Conditional Branch Instruction
5. Unconditional jump Instruction

17.Draw the basic structure of Memory Hierarchy.[APR 2019]

18. What you meant by transistors?

A transistor is simply an on/off switch controlled by electricity. The *integrated circuit* (IC) combined dozens to hundreds of transistors into a single chip.

19. What you meant by die?

The individual rectangular sections that are cut from a wafer, more informally known as chips.

20. Define the terms response time and throughput.

The response time is the time between the start and completion of a task also referred to as execution time. The throughput or bandwidth the total amount of work done in a given unit of time.

21. How will you maximize the performance of computer?

To maximize performance, we want to minimize response time or execution time for some task. Thus, we can relate performance and execution time for a computer X:

22. Can you compare the performance of two different computers quantitatively?

Yes, we can relate the performance of two different computers quantitatively as Below Then we could say the X is *n* times faster than Y.

23. If computer A runs a program in 10 seconds and computer B runs the same program in 15 seconds how much faster is A than B?

We know that A is *n* times as fast as B if

$$\frac{\text{Performance}_A}{\text{Performance}_B} = \frac{\text{Execution time}_B}{\text{Execution time}_A} = n$$

Thus the performance ratio is

$$\frac{15}{10} = 1.5$$

and A is therefore 1.5 times as fast as B. In the above example, we could also say that computer B is 1.5 times *slower than* computer A, since

$$\frac{\text{Performance}_A}{\text{Performance}_B} = 1.5$$

means that

$$\frac{\text{Performance}_A}{1.5} = \text{Performance}_B$$

24. Compare user CPU time and system CPU time.

user CPU time system CPU time The CPU time spent in a program itself The CPU time spent in the operating system performing tasks on behalf of the program

25. What is the difference between uniprocessor and multiprocessor?

Uniprocessor: - A type of architecture that is based on a single computing unit. All operations (additions, multiplications, etc) are done sequentially on the unit. Multiprocessor: - A type of architecture that is based on multiple computing units. Some of the operations (not all, mind you) are done in parallel and the results are joined afterwards.

26. Define MIPS.

MIPS is an instruction execution rate, MIPS specifies performance inversely to execution time; faster computers have a higher MIPS rating.

27. Define the Term WORD.

The natural unit of access in a computer, usually a group of 32 bits; corresponds to the size of a register in the MIPS architecture.

28. What are called data transfer instructions?

MIPS must include instructions that transfer data between memory and registers. Such instructions are called data transfer instructions. To access a word in memory, the instruction must supply the memory address. Memory is just a large, single-dimensional array, with the address acting as the index to that array, starting at 0.

29.State Amdahl's Law.(APR 2019)

Amdahl's law is a formula used to find the maximum improvement improvement possible by improving a particular part of a system. In parallel computing,

Amdahl's law is mainly used to predict the theoretical maximum speedup for program processing using multiple processors.

$$\text{Speedup} = \frac{\text{Performance for entire task using the enhancement when possible}}{\text{Performance for entire task without using the enhancement}}$$

Alternatively,

$$\text{Speedup} = \frac{\text{Execution time for entire task without using the enhancement}}{\text{Execution time for entire task using the enhancement when possible}}$$

30. Which is called alignment restriction?

In MIPS, words must start at addresses that are multiples of 4. This requirement is called an alignment restriction.

31.List out the methods used to improve system performance.

The methods used to improve system performance are

1. Processor clock
2. Basic Performance Equation
3. Pipelining
4. Clock rate
5. Instruction set
6. Compiler

32.Define response time

Also called **execution time**. The total time required for the computer to complete a task, including disk accesses, memory accesses, I/O activities, operating system overhead, CPU execution time, and so on.

33.What is Throughput?

Also called **bandwidth**. Another measure of performance, it is the number of tasks completed per unit time.

34. Give the format of MIPS R-type instruction.

op: Basic operation of the instruction, traditionally called the opcode.
rs: The first register source operand.
rt: The second register source operand.
rd: The register destination operand. It gets the result of the operation.
shamt: Shift amount.
funct: Function. This field, often called the *function code*, selects the specific variant of the operation in the op field.

UNIT II - ARITHMETIC OPERATIONS

1. Define Arithmetic Logic Unit (ALU).

Hardware that performs addition, subtraction, and usually logical operations such as AND and OR.

2. When can overflow occur in addition?

Overflow occurs when adding two positive numbers and the sum is negative, or vice versa. This spurious sum means a carry out occurred into the sign bit.

3. When the overflow will not occur in addition?

When adding operands with different signs, overflow cannot occur. The reason is the sum must be no larger than one of the operands. For example, −10+4=−6. Since the operands fit in 32 bits and the sum is no larger than an operand, the sum must fit in 32 bits as well. Therefore, no overflow can occur when adding positive and negative operands.

4. Write the principle of booth multiplication.

Booth multiplication is nothing but addition of properly shifted multiplicand patterns.

It is carried out by following steps:

a) Start from LSB. Check each bit one by one.

b) Change the first one as -1.

c) Skip all exceeding one‟s (record them as zeros) till you see a zero. Change this zero as one

d) Continue to look for next one without disturbing zeros, precede using rules b), and c)

5. List the advantages of Booth algorithm.

1. It handles both positive and negative multipliers uniformly.
2. It achieves some efficiency in the no. of additions required When the multiplier has a few large blocks of 1‟s.

6.What are the main features of Booth's algorithm?

1. It handles both positive and negative multipliers uniformly.
2. It achieves some efficiency in the number of addition required when

the multiplier has a few large blocks of 1s.

7. What you meant by overflow and underflow in Floating point numbers?

overflow means that the exponent is too large to be represented in the exponent field.

overflow (floating-point)

A situation in which a positive exponent becomes too large to fit in the exponent field.

underflow (floating-point)

A situation in which a negative exponent becomes too large to fit in the exponent field.

8. What are the floating point instructions in MIPS?

MIPS supports the IEEE 754 single precision and double precision formats with these instructions:

- Floating-point addition
- Floating-point subtraction
- Floating-point multiplication
- Floating-point division
- Floating-point comparison
- Floating-point branch

9. Define ULP

Units in the Last Place is defined as the number of bits in error in the least significant bits of the significant between the actual number and the number that can be represented.

10. What is meant by sticky bit?

Sticky bit is a bit used in rounding in addition to guard and round that is set whenever there are nonzero bits to the right of the round bit. This sticky bit allows the computer to see the difference between 0.50 … 00 ten and .... 01 ten when rounding.

11. What is meant by sub-word parallelism?

Given that the parallelism occurs within a wide word, the extensions are classified as subword parallelism. It is also classified under the more general name of data level parallelism. They have been also called vector or SIMD, for single instruction, multiple data.

By partitioning the carry chains within a 128-bit adder, a processor could use Parallelism to perform simultaneous operations on short vectors of sixteen 8-

bit operands, eight 16-bit operands, four 32-bit operands, or two 64-bit operands. The cost of such partitioned adders was small. Given the parallelism occurs within a wide word, the extensions are classified as *sub word parallelism*.

12. What are the steps in the floating-point addition?

The steps in the floating-point addition are

1. Align the decimal point of the number that has the smaller exponent.
2. Addition of the significant numbers.
3. Normalize the sum.
4. Round the result

13. Define Guard and Round

Guard is the first of two extra bits kept on the right during intermediate calculations of floating point numbers. It is used to improve rounding accuracy. Round is a method to make the intermediate floating-point result fit the floating point format; the goal is typically to find the nearest number that can be represented in the format. IEEE 754, therefore, always keeps two extra bits on the right during intermediate additions, called guard and round, respectively.

14. Define ULP

Units in the Last Place is defined as the number of bits in error in the least significant bits of the significant between the actual number and the number that can be represented.

15. Write the Add/subtract rule for floating point numbers.

1) Choose the number with the smaller exponent and shift its mantissa right a number of steps equal to the difference in exponents.
2) Set the exponent of the result equal to the larger exponent.
3) Perform addition/subtraction on the mantissa and determine the sign of the result
4) Normalize the resulting value, if necessary.

16. Write the multiply rule for floating point numbers.

1) Add the exponent and subtract 127.
2) Multiply the mantissa and determine the sign of the result.
3) Normalize the resulting value, if necessary.

24. What is the purpose of guard bits used in floating point arithmetic

Although the mantissa of initial operands are limited to 24 bits, it is important to retain extra bits, called as guard bits.

17. What are the ways to truncate the guard bits?

There are several ways to truncate the guard bits:

1) Chopping
2) Von Neumann rounding
3) Rounding

18. Define carry save addition(CSA) process.

Instead of letting the carries ripple along the rows, they can be saved and introduced into the next roe at the correct weighted position. Delay in CSA is less than delay through the ripple carry adder.

19.Define chopping.

Chopping is a simple way to truncate or remove the guard bits and make no changes in the retained bits.

20.Define Half adder and full adder.

The logic circuit that performs the addition of two binary digits is known as half adder. The circuit that performs the addition of three binary digits is known as full adder.

21.What is excess-127 format?

Instead of the signed exponent E, the value actually stored in the exponent field is and unsigned integer E. In some cases, the binary point is variable and is automatically adjusted as computation proceeds. In such case, the binary point is said to float and the numbers are called floating point numbers.

22.Define Interrupt

An exception that comes from outside of the processor. (Some architectures use the term *interrupt* for all exceptions.)

23.What is bit pair recoding? Give an example.

Bit pair recoding halves the maximum number of summands. Group the Booth- recoded multiplier bits in pairs and observe the following: The pair (+1 -1) is equivalent to to the pair (0 +1). That is instead of adding -1 times the multiplicand m at shift position i to +1 M at position i+1, the same result is obtained by adding +1 M at position i.

Eg: 11010 – Bit Pair recoding value is 0 -1 -2.

24.Define Single precision

A floating-point value represented in a single 32-bit word.

1. s is the sign of the floating-point number (1 meaning negative),

2. exponent is the value of the 8-bit exponent field (including the sign of the exponent), and

3. *fraction* is the 23-bit number.

25.State double precision floating point number?

Double-precision floating-point format is a computer **number** format that occupies 8 bytes (64 bits) in computer**,** computer memory and represents a wide, dynamic range of values by using a **floating point.**

UNIT III - PROCESSOR AND CONTROL UNIT

PART A – QUESTIONS

1. Define hazard and its types.

Any condition that causes the pipeline to stall is called a hazard. Its types are:

1. Data hazard
2. Instruction hazard
3. Structural hazard

2. Define data hazard.

1. A data hazard is any condition in which either the source or the destination operands of an instruction are not available at the time expected in the pipeline.
2. A data hazard is a situation in which the pipeline is stalled because the data to be operated on are delayed for some reason.

3. Define instruction hazard or control hazard.

A pipeline may also be stalled because of the delayed in the availability of an instruction. This may be a result of a miss in the catch, requiring the instruction to be fetched from the main memory. Such hazard are often called control hazard.

4. Define structural hazard.

Structural hazard occurs in the following situation when two instructions required the use of a given hardware resource at the same time.

5. Define operand forwarding.

The data are available at the output of the ALU once the Execute stage completes step E1. Hence, the delay can be reduced or possibly eliminated if we arrange for the result of instruction I1 to be forwarded directly for use in step E2. This is called operand forwarding.

6. List the five stages of instruction execution.

1. IF: Instruction fetch
2. ID: Instruction decode and register file read
3. EX: Execution or address calculation
4. MEM: Data memory access
5. WB: Write back

7. Write about the classification of data hazards.

Classification of data hazard: A pair of instructions can produce data hazard by referring reading or writing the same memory location. Assume that i is executed before J. So, the hazards can be classified

1. RAW hazard
2. WAW hazard
3. WAR hazard

8. Define Pipelining.

In order to reduce the overall processing time several instructions are being executed simultaneously. This process is termed as pipelining.

9. How data hazard can be prevented in pipelining?

Data hazards in the instruction pipelining can prevented by the following techniques.

a) Operand Forwarding
b) Software Approach

10. Define Superscalar processor.

There are processors which are capable of achieving an instruction executing throughput of more than one instruction per cycle. They are known superscalar processor.

11. List out the methods used to improve system performance.

The methods used to improve system performance are

1. Processor clock
2. Basic Performance Equation
3. Pipelining
4. Clock rate
5. Instruction set
6. Compiler

12. What is branch folding?

When the instruction fetch unit executes the branch instruction concurrently with the execution of the other instruction, then this technique is called branch folding.

13. Define exception and interrupt.

Exception:

The term exception is used to refer to any event that causes an interruption.

Interrupt:

An exception that comes from outside of the processor. There are two types of interrupt.

1. Imprecise interrupt and 2. Precise interrupt

14. What is branch Target Address?

The address specified in a branch, which becomes the new program counter, if the branch is taken. In MIPS the branch target address is given by the sum of the offset field of the instruction and the address of the instruction following the branch.

15. Classify the pipeline computers.

Based on level of processing → processor pipeline, instruction pipeline, arithmetic pipelines

Based on number of functions→ Uni-functional and multi-functional pipelines.

Based on the configuration → Static and Dynamic pipelines and linear and non linear pipelines

Based on type of input→ Scalar and vector pipelines.

16. Write down the expression for speedup factor in a pipelined architecture.

The speedup for a pipeline computer is S = (k + n -1) tp Where, K → number of segments in a pipeline,N → number of instructions to be executed. Tp → cycle time

$$\frac{\text{Time per instruction on unpipelined machine}}{\text{Number of pipe stages}}$$

17. What are the problems faced in instruction pipeline.

Resource conflicts → Caused by access to the memory by two at the same time. Most of the conflicts can be resolved by using separate instruction and data memories.

Data dependency → Arises when an instruction depends on the results of the previous instruction but this result is not yet available.

Branch difficulties → Arises from branch and other instruction that change the value of PC (Program Counter).

18. Define Sign-extend in data path.

To increase the size of a data item by replicating the high-order sign bit of the original data item in the high-order bits of the larger, destination data item. a unit to sign-extend the 16-bit offset field in the instruction to a 32-bit signed value

19. What is Delayed branch?

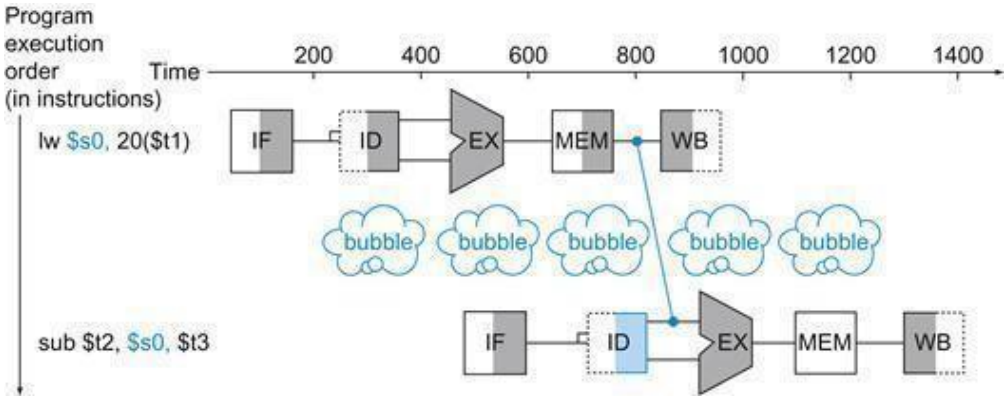
A type of branch where the instruction immediately following the branch is always executed, independent of whether the branch condition is true or false.

20. What are the control lines of MIPS functions.

ALU control lines	Function
0000	AND
0001	OR
0010	add
0110	Sub

21. Define Pipeline stall

Pipeline stall is also called as bubble. A stall initiated in order to resolve a hazard.



22. What are the Schemes for resolving control hazards?

1. Assume Branch Not Taken
2. Reducing the Delay of Branches
3. Dynamic Branch Prediction

23. What is Control Hazard?

Control hazard is also called as **branch hazard**. When the proper instruction cannot execute in the proper pipeline clock cycle because the instruction that was fetched is not the one that is needed; that is, the flow of instruction addresses is not what the pipeline expected.

24. Name control signal to perform arithmetic operation.

1.Regdst
2.Regwrite
3.ALU Src

25. Define Correlating, Tournament branch predictor

Correlating predictor

A branch predictor that combines local behavior of a particular branch and global information about the behavior of some recent number of executed branches.

Tournament branch predictor

A branch predictor with multiple predictions for each branch and a selection mechanism that chooses which predictor to enable for a given branch

UNIT 4 – PARALLELISM

PART – A QUESTIONS

1. What is Instruction level parallelism?

ILP is a measure of how many of the operations in a computer program can be performed simultaneously. The potential overlap among instructions is called instruction level parallelism

2. What are the various types of Dependences in ILP.

1. Data Dependences
2. Name Dependences
3. Control Dependences

3. Define Static multiple issue and Dynamic multiple issue.

Static multiple issue -An approach to implementing a multiple-issue processor where many decisions are made by the compiler before execution.

Dynamic multiple issue -An approach to implementing a multiple issue processor where many decisions are made during execution by the processor.

4. Define Reorder buffer?

The buffer that holds results in a dynamically scheduled processor until it is safe to store the results to memory or a register.

5. Define Out of order execution.

A situation in pipelined execution when an instruction blocked from executing does not cause the following instructions to wait.

6. Differentiate UMA from NUMA.

Uniform memory access (UMA)is a multiprocessor in which latency to any word in main memory is about the same no matter which processor requests the access.

Non uniform memory access (NUMA) is a type of single address space multiprocessor in which some memory accesses are much faster than others depending on which processor asks for which word.

7. What are the advantages of Speculation?

1. Speculating on certain instructions may introduce exceptions that were formerly not present.
2. Example a load instruction is moved in a speculative manner, but the address it uses is not legal when the speculation is incorrect.

3. Compiler-based speculation, such problems are avoided by adding special speculation support that allows such exceptions to be ignored until it is clear that they really should occur.

8. Write down the difference between this **simple superscalar** and a **VLIW processor**

1. The code, whether scheduled or not, is guaranteed by the hardware to execute correctly.
2. The compiled code always run correctly independent of the issue rate or pipeline structure of the processor.
3. In some VLIW designs, recompilation was required when moving across different processor models.

9. What are all the Advantages SMT.

1. Simultaneous Multithreaded Architecture is superior in performance to a multiple-issue multiprocessor (multiple-issue CMP).
2. SMP boosts utilization by dynamically scheduling functional units among multiple threads.
3. SMT also increases hardware design flexibility.
4. SMT increases the complexity of instruction scheduling.

10. What are all the three major units of Dynamic pipeline scheduling?

1. instruction fetch and issue unit
2. multiple functional units
3. commit unit

11. Define VLIW.

A style of instruction set architecture that launches many operations that are defined to be independent in a single wide instruction, typically with many separate opcode fields.

12. What you meant by register renaming?

As instructions are issued, the register specifiers for pending operands are renamed to the names of the reservation station in a process called **register renaming**. This combination of issue logic and reservation stations provides renaming and eliminates WAW and WAR hazards.

13. Consider a nonpipelined machine with 6 execution stages of lengths: 50 ns, 50 ns, 60 ns, 60 ns, 50 ns, and 50 ns.
- Find the instruction latency on this machine.
- How much time does it take to execute 100 instructions?

Instruction latency = 50+50+60+60+50+50= 320 ns
Time to execute 100 instructions = 100*320 = 32000 ns

14. What is Flynn's Classification?

1. In 1966, Michael Flynn proposed a classification for computer architectures based on the number of instruction steams and data streams (Flynn‘s Taxonomy).
2. Flynn uses the stream concept for describing a machine's structure A stream simply means a sequence of items (data or instructions).
3. The classification of computer architectures based on the number of instruction steams and data streams (Flynn‘s Taxonomy).

15. Describe about SIMD.

SIMD (Single-Instruction stream, Multiple-Data streams) Each instruction is executed on a different set of data by different processors i.e multiple processing units of the same type process on multiple-data streams.

This group is dedicated to array processing machines. Sometimes, vector processors can also be seen as a part of this group.

16. Define – Issue Slots and Issue Packet

Issue slots are the positions from which instructions could be issued in a given clock cycle.

By analogy, these correspond to positions at the starting blocks for a sprint. Issue packet is the set of instructions that issues together in one clock cycle; the packet may be determined statically by the compiler or dynamically by the processor.

17. What is SMT?

Simultaneous Multithreading (SMT)is a variation on hardware multithreading that uses the resources of a multiple-issue, dynamically scheduled pipelined processor to exploit thread level parallelism. It also exploits instruction level parallelism.

18. What are the three multithreading options?

The three multithreading options are:
1. A superscalar with coarse-grained multithreading
2. A superscalar with fine-grained multithreading
3. A superscalar with simultaneous multithreading

19. What is meant by loop unrolling?

An important compiler technique to get more performance from loops is loop unrolling, where multiple copies of the loop body are made. After unrolling, there is more ILP available by overlapping instructions from different iterations.

20. What are the characteristics of SRAMS.

1. SRAMS are fast

2. They are volatile
3. They are of high cost
4. Less density

21. what are asynchronous DRAMS.

In asynchronous DRAMS, the timing of the memory device is controlled asynchronously. A specialized memory controller circuit provides the necessary control signals RAS and CAS that govern the timing. The processor must take into account the delay in the response of the memory.

22. What are synchronous DRAMS.

Synchronous DRAMS are those whose operation is directly synchronized with a clock signal.

23. Name the common replacement algorithms.

1. Least-Recently Used(LRU)
2. First In First Out(FIFO)
3. Least Frequently Used(LFU)
4. Random