

GCE A LEVEL – COMPUTER SCIENCE UNIT 4 QUESTION PACK

1500U40-1 · 2015 spec Unit 4 Topic 8 · A2 unit, first sat 2017, 100 marks, 2h paper

REVISE.wales

COMPUTER SCIENCE – UNIT 4 · File Organisation – Direct, Sequential, Indexed & Hashing

Topic 4.8 – File-management methods including direct (hashed), sequential, indexed-sequential and multi-level indexes

Comparing direct (hashed) access, sequential and indexed-sequential file organisations; calculating hash addresses and handling collisions via overflow areas; reading and following multi-level indexes; choosing fixed-length vs variable-length records; and explaining the role of a file-manager program.

2015 specification · current

Estimated time for entire question pack: ~58 min

Derived from the Unit 4 pace of ~1.5 min/mark, padded for written-prose answers (39 marks over 6 questions).

*You are advised to **not** attempt to complete all of this in one sitting.*

ABOUT THIS QUESTION PACK

This is a **comprehensive topic question pack**, not a single mock paper. It contains every question from the WJEC A2 Unit 4 papers (Summer 2017 – Summer 2024, COVID gap) that maps onto Topic 4.8 of the 2015 specification.

Questions are ordered by source paper date.

INSTRUCTIONS

Use black ink or black ball-point pen. Show all working. A calculator is allowed where useful.

All question content is © WJEC CBAC Ltd. and reproduced for revision purposes.

For Examiner's use only

Q	Source	Max	Mark
1	S17 Q11	6	
2	S19 Q6	7	
3	S19 Q7	5	

Q	Source	Max	Mark
4	S22 Q1	4	
5	S23 Q7	11	
6	S24 Q2	6	
Total		39	

File Organisation – Direct, Sequential, Indexed & Hashing – what the spec asks

WJEC GCE A Level Computer Science (from 2015) · Unit 4: Computer Architecture, Data, Communication & Applications · Topic 4.8.

Direct (hashed) access

- Apply hash function to key → storage location.
- Typically $\text{address} = \text{key} \text{ MOD } \text{size}$.
- Constant-time $O(1)$ access on average.
- Choose load factor < 1 to keep collisions rare.

Sequential access

- Records stored one after another in physical order.
- Insertion is expensive (shift records) unless append-only.
- Excellent for processing the whole file in order (e.g. payroll).
- Searching for one record is $O(n)$.

Indexed-sequential & multi-level indexes

- Records held sequentially, plus an index of (key, address) pairs.
- Multi-level: top-level index points into a second-level index → cuts disk I/O for huge files.
- Search top index by binary / hashed lookup, follow the chain down.
- Inserts: write to overflow area; periodic reorganisation merges in.

Overflow areas

- When a hash address is already taken, the colliding record goes to an overflow region.
- On retrieval, check the primary location, then walk the overflow chain.
- Performance degrades as the overflow grows – trigger reorganisation when load factor is too high.
- Choosing a hashing algorithm that spreads keys evenly minimises overflow use.

Fixed vs variable length records

- Fixed: all records the same size – direct indexing by $\text{start} + n \times \text{size}$.
- Variable: each record is preceded by its length, or a delimiter separates fields.
- Fixed wastes space for short records but is fast.
- Variable saves space (e.g. words in a dictionary) but slows random access.

File-manager program

- Provides a user interface to the underlying file system.
- Common operations: create, copy, move, rename, delete, search, change permissions.
- Organises files into a tree of folders / directories.
- Translates user actions into the OS calls that touch the disk.

File Organisation – Direct, Sequential, Indexed & Hashing in one page

Quick-reference notes – revisit before each question.

Hash address

address = key MOD size.
Constant-time average.
Collisions → overflow area.
Choose a hash that spreads keys evenly.

Indexed-sequential

Sequential data file + index of (key, addr).
Multi-level index: index of indexes for huge files.
Inserts → overflow, reorganised periodically.

Direct vs sequential

Direct: random access, fast for any one record.
Sequential: cheap for full-file passes (payroll).
Indexed-seq: best of both.

Overflow handling

Primary slot occupied → spill to overflow.
On read, follow the chain.
Reorg merges overflow back when load factor is high.

Fixed vs variable

Fixed: $\text{addr} = \text{start} + n \times \text{size}$; fast random.
Variable: length prefix or delimiter; saves space.
Database rows: usually variable, with row pointers.

File-manager UI

Browse, copy, move, rename, delete, search, permissions.
Folders form a tree.
Translates user actions into OS calls.

10. Expert systems are widely used by organisations for a variety of purposes. Describe the benefits to an organisation of using an expert system. [8]

11. Explain the use of multi-level indexes and draw a diagram to demonstrate the operation of a three-level index. [6]

12. Khan's Pharmaceuticals currently uses an ID card system to control employee access to its premises. This has proved problematical with employees swapping cards and the company now wishes to use a voice print recognition system in its place.

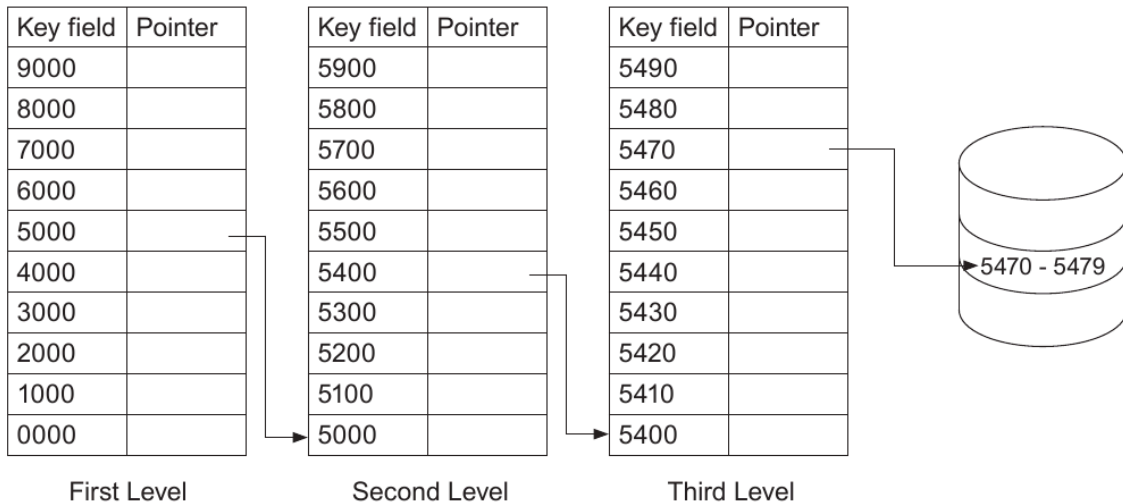
Describe how this system would operate and explain the benefits and drawbacks associated with a biometric system used for this purpose. [11]

END OF PAPER

6. A three-level indexed sequential file system stores customer records which have key field values in the range 0000 to 9999. Only a few of the possible key field values have currently been allocated to customers.

To save memory space, index blocks are only created when required. At present, only the three index blocks shown in the diagram have been created.

Customer records are stored on disk in data blocks. Each data block may contain up to 10 customer records. Records within any one data block have key field values which differ only in the last digit, e.g. 5470 to 5479, and are sorted sequentially.



- (a) Explain how new customer records 3892 and 3893 are to be added to the system, mentioning the creation and linking of any necessary index or data blocks. [5]
- (b) Explain how the computer would access record 3893 at a later date. [3]
7. A sequential list of English language words can be stored either in fixed length or variable length records. Six thousand (6000) words are to be stored. The maximum length of any word is 22 letters, with an average word length of 7 letters.

Variable length records have a format where each word is terminated with an @ character (which requires one byte of storage):

Estimate the sizes of the files required to store the words in:

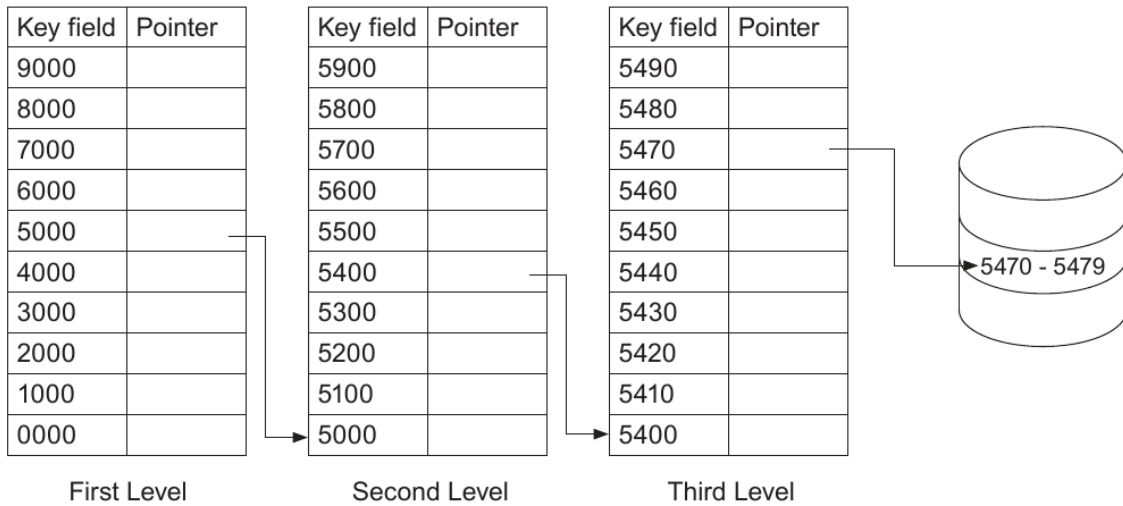
- (a) fixed length format
- (b) variable length format.

Give your answers in kilobytes, correct to one decimal place. [4]

6. A three-level indexed sequential file system stores customer records which have key field values in the range 0000 to 9999. Only a few of the possible key field values have currently been allocated to customers.

To save memory space, index blocks are only created when required. At present, only the three index blocks shown in the diagram have been created.

Customer records are stored on disk in data blocks. Each data block may contain up to 10 customer records. Records within any one data block have key field values which differ only in the last digit, e.g. 5470 to 5479, and are sorted sequentially.



- (a) Explain how new customer records 3892 and 3893 are to be added to the system, mentioning the creation and linking of any necessary index or data blocks. [5]
- (b) Explain how the computer would access record 3893 at a later date. [3]

7. A sequential list of English language words can be stored either in fixed length or variable length records. Six thousand (6000) words are to be stored. The maximum length of any word is 22 letters, with an average word length of 7 letters.

Variable length records have a format where each word is terminated with an @ character (which requires one byte of storage):

Estimate the sizes of the files required to store the words in:

- (a) fixed length format
- (b) variable length format.

Give your answers in kilobytes, correct to one decimal place. [4]

8. Define the term **data mining** and describe how three different organisations might use data mining. [8]
9. Describe four data validation techniques. [4]
10. (a) Describe the types of malicious software which might be transferred to computers and the delivery mechanisms, and the steps that can be taken to protect against these. [6]
- (b) Computer data may be at its highest security risk during transfer from one location to another. Outline the risks that exist at this time, and how they can be minimised. [4]
11. A city is developing a new light railway system to connect the city centre to the surrounding suburbs. The system will use driverless trains, and several computer centres will control different areas of the network.
- (a) Give examples of input and output which might be required by control systems on board the trains. [3]
- (b) The system will be safety critical. Explain what is meant by a safety critical system, describing measures that are involved in ensuring safety. [5]
12. Describe the operation of a mainframe computer using a multi-programming, multi-user operating system. [7]
13. Explain what is meant by distributed processing, and describe how this will operate using an example that you have studied. [6]

END OF PAPER

Answer **all** questions.

1. A file manager program provides an interface to a computer's file system.
 - (a) Describe how files may be organised and managed using a file manager program. [3]
 - (b) Files may be organised using a hashing algorithm. Explain the purpose of a hashing algorithm. [3]

2. Interrupts are often generated in a computer system.
 - (a) Identify a situation in which a high-priority interrupt may be generated. [1]
 - (b) Describe the procedure for dealing with an interrupt with a higher priority than the interrupt that is currently being processed. [3]

3. An assembly language program is required to check a four digit numeric code against a code stored in location Y.

If the code is entered correctly then the program should output 1. If the code entered is incorrect the program should output 2 and should allow the user to re-enter a code.

The assembly language used by the processor has an instruction set which includes the following commands:

Assembly Language Command	Description
INP	Input numeric value to the accumulator
OUT X	Output a numeric value X
STA X	Copy contents of accumulator to memory location X
LDA X	Load accumulator with contents of memory location X
HLT	End
SUB Y	Subtract the numeric value in location Y from the accumulator, leaving the result in the accumulator
JZE LABEL	Jump to LABEL if the contents of the accumulator are equal to zero
JMP LABEL	Jump unconditionally to LABEL

Using the instruction set, write a program to check the code entered. [6]

7. The owner of a shop requires a program to store stock records. Each stock item has a three-digit identification code in the range 000 to 999. The number of different stock items held by the shop is currently 95.

A programmer considers two different methods of storing the stock records:

- A direct access file with 100 storage locations. The location for each item is calculated using the hash function: $\text{identificationCode} \text{ MOD } 100$.
- A sequential access file with 1000 storage locations. The stock record is stored at the location corresponding to its identification code, so product 715 will be stored at location 715.

Explain the advantages and disadvantages of these two methods. [6]

8. Describe the three basic states of a process in a multi-tasking computer. [6]

9. Improved worker safety is a benefit of automation.

(a) Describe **two other** benefits and **two** drawbacks of automating a manufacturing process. [4]

(b) Explain, giving an example of each type of system, the main purpose of:

- a safety related control system
- a safety critical control system.

[4]

Answer **all** questions.

1. Describe the role of three of the main components of a contemporary Central Processing Unit (CPU). [6]

2. AJ Jewellers uses a direct access file with a separate overflow area to store stock records. Each record has a key field made up from a four-digit code followed by the year number, for example: 137524, 701924.

It is proposed to use the hashing algorithm: key field MOD 1000

- (a) Calculate the location of the stock records for each of the two example key fields. [2]

 - (b) Explain why this hashing algorithm is unsuitable. [3]

 - (c) Give an example of a more suitable hashing algorithm for the file. [1]
-
3. AJ Jewellers has invested in a 3D printing system to produce objects in precious metals. The system includes CAD software and a double buffering printer interface.

Explain why the 3D printing system uses double buffering. [4]

END OF QUESTION PACK

6 questions · 39 marks · ~58 min

Source: WJEC A2 Computer Science Unit 4 (1500U40-1), Summer 2017–2024, COVID gap
Curated for WJEC Computer Science 2015 spec A2 Unit 4 – Topic 8 (4.8)

© WJEC CBAC Ltd. Pack layout © revise.wales for revision purposes only.