File Processing

Agenda

Introduction

File and Streams

Creating a Sequential File

Reading Data From a Sequential File

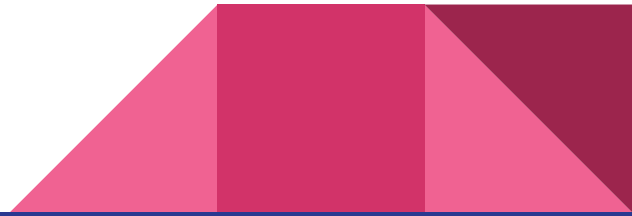
Random Access File

Reading Data From Random Access File

Writing Data Randomly to a Random Access File

Reading from a Random Access File Sequnetially

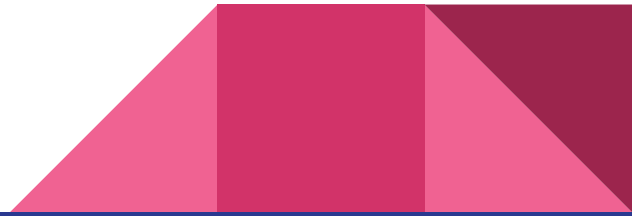
Object Serialization



Usage of Files

Purpose of using Files - Data Persistence - permanent retention of data

Files in computers are stored on a secondary storage - hard disks,CDs, DVDs, flash disks and tapes



Files and Streams

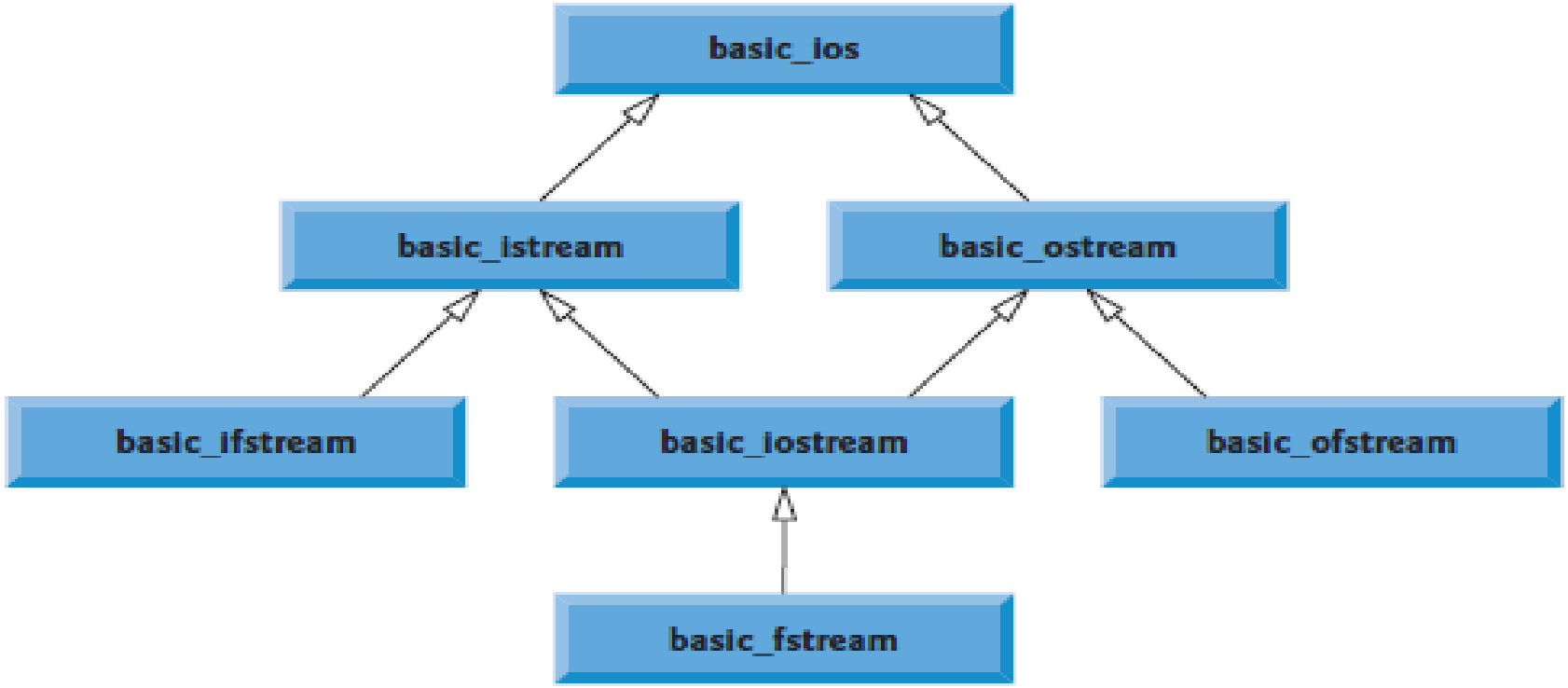
- Files in C++ are viewed as Sequence of bytes

- Each file either ends with an end-of-file marker, or at a specific byte number recorded in operating system maintained administrative data structure.

- When a file is opened, an object is created, and a stream is associated with an object.

- The streams associated with these objects provide communication channels between a program and a particular file or device.

File Processing Class Templates



Creating a Sequential File

- In C++, a file has no structure.
- Programmer needs to structure the file, to meet the application's requirements.
- A Sequential file, used in an accounts-receivable system to help manage the money own to a company by it's credit clients, will have client's account number, name and balance.
- This data for each client will be 'record' for each client. The account number serves as the record key;

Reading Data From a Sequential File

- Files store data so it may be retrieved for processing when needed.
- Creating an ifstream object opens a file for input.
- The ifstream constructor can receive the filename and the file open mode as arguments
- Opens the file and establishes a "line of communication" with the file.

Random Access File

- Sequential files are inappropriate for instant-access applications, to locate a particular record immediately
- Application such as airlines reservation systems, banking systems etc, needs rapid access to specific data.
- By creating Random Access File, a particular information can be located instantly.
- Individual records of a random-access file can be accessed directly (and quickly) without having to search other records.

Creating Random Access File

- To create a Random Access File, a variety of techniques can be used.
- Easiest method is to require that all records in a file be of the same fixed length.
- By Using same-size, fixed-length records makes it easy for a program to quickly calculate (as a function of the record size and the record key) the exact location of any record relative to the beginning of the file.

Credit Processing Program

Consider the following problem statement:

Create a credit-processing program capable of storing at most 100 fixed-length records for a company that can have up to 100 customers. Each record should consist of an account number that acts as the record key, a last name, a first name and a balance. The program should be able

-to update an account

- insert a new account,

- delete an account

- insert all the account records into a formatted text file for printing.

Object Serialization

- In a disk file, the object's type information is not stored, but only the attributes.
- Knowing the object type to which the data corresponds makes it easier to read the data into an object of that type.
- However, how to distinguish, if object of different types are stored in the same file?

- One approach to this problem is known as Serialization of object.
- Serialized object is an object represented as a sequence of bytes that includes the object's data as well as information about the object's type and the types of data stored in the object
- C++ does not provide built-in serialization mechanism, but by using third party and open source C++ libraries, it can be achieved.
- The most famous library used for this purpose is Boost Library algorithm.

Fill in the Blanks

a) Member function _____ of the file streams fstream, ifstream and ofstream closes a file.

b) The istream member function _____ reads a character from the specified stream.

c) Member function _____ of the file streams fstream, ifstream and ofstream opens a file.

d) The istream member function _____ is normally used when reading data from a file in random-access applications.

e) Member functions _____ and _____ of istream and ostream set the file-position pointer to a specific location in an input or output stream, respectively.

Ans) a) close. b) get. c) open. d) read. e) seekg, seekp.

True or False

a) Member function read cannot be used to read data from the input object cin.

b) You must create the cin, cout, cerr and clog objects explicitly.

c) A program must call function close explicitly to close a file associated with an ifstream, ofstream or fstream object.

d) If the file-position pointer points to a location in a sequential file other than the beginning

of the file, the file must be closed and reopened to read from the beginning of the file.

e) The ostream member function write can write to standard-output stream cout.

f) Data in sequential files always is updated without overwriting nearby data.

g) Searching all records in a random-access file to find a specific record is unnecessary.

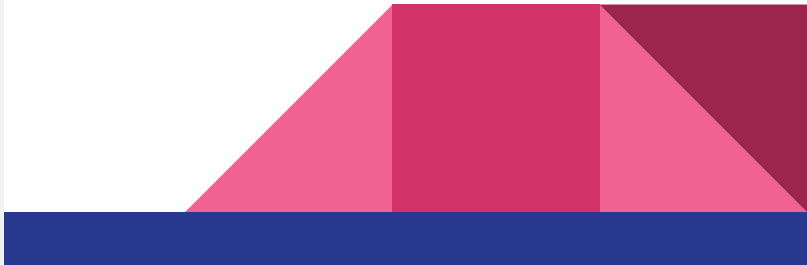
h) Records in random-access files must be of uniform length.

i) Member functions seekp and seekg must seek relative to the beginning of a file.

Programming Exercises

1> (Hardware Inventory) You are the owner of a hardware store and need to keep an inventory that can tell you what different tools you have, how many of each you have on hand and the cost of each one. Write a program that initializes the random-access file hardware.dat to 100 empty records, lets you input the data concerning each tool, enables you to list all your tools, lets you delete a record for a tool that you no longer have and lets you update any information in the file. The tool identification number should be the record number. Use the following information to start your file:

Record #	Tool name	Quantity	Cost
3	Electric sander	7	57.98
17	Hammer	76	11.99
24	Jig saw	21	11.00
39	Lawn mower	3	79.50
56	Power saw	18	99.99
68	Screwdriver	106	6.99
77	Sledge hammer	11	21.50
83	Wrench	34	7.50



2> (sizeof Operator) Write a program that uses the sizeof operator to determine the sizes in bytes of the various data types on your computer system. Write the results to the file datasize. dat, so that you may print the results later. The results should be displayed in two-column format with the type name in the left column and the size of the type in right column, as in:

```
char                1
unsigned char       1
short int           2
unsigned short int  2
int                 4
unsigned int        4
long int            4
unsigned long int   4
float               4
double              8
long double         10
```