1. Answer the following questions:

i) Why must the meta data be established for a database?

A fundamental characteristic of the database approach is that the database system contains not only the database itself but also a complete definition or description of the database structure and constraints. This definition is stored in the system **catalog,** which contains information such as the structure of each file, the type and storage format of each data item, and various constraints on the data. The information stored in the catalog is called **meta-data,** and it describes the structure of the primary database

The catalog is used by the DBMS software and also by database users who need information about the database structure. A general purpose DBMS software package is not written for a specific database application, and hence it must refer to the catalog to know the structure of the files in a specific database, such as the type and format of data it will access.

Meta data is kind of data that is stored in the system catalog. It is a set of relations. Data from the database table are described as meta data in the system catalog. Meta data is very important to be established for a database. Because, system catalog is closely related to the database and system catalog contains all the meta data. When we try to explore this system catalog that is known to be the heart of the any general purpose database management system (DBMS). As containing the meta data system catalog considered as a mini database itself. And its main one of the main functions is to store the schemas or descriptions of the database that DBMS contains. And all these important information are called meta data. Meta data is important because it includes a description of the conceptual database schema, the internal schema, and external schemas, and the mappings between schemas at different levels. Also, information that are needed by specific DBMS modules such as, the query optimization module or the security and authorization module is stored in the catalog.

ii) What is the meaning of each attribute in the relation schema shown below?

RELATION_INDEXES

REL_NAME	INDEX_NAME	MEMBER_ATTR	INDEX_TYPE	ATTR_NO	ASC_DESC

Use an example to explain your answer.

Works_on	I1	SSN	Primary	1	ASC
Works_on	I1	Pno	Primary	2	ASC
Works_on	I2	SSN	Clustering	1	ASC

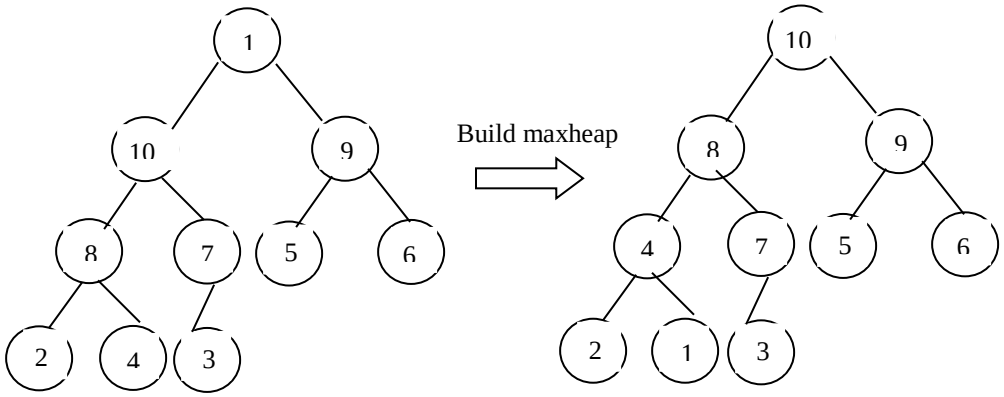
The Meaning of each attribute in the relational schema are as follows:

Relational name indicates the name of the relation, for example, create an index for relation “WORKS_ON” will have index name “Works_on”. Each

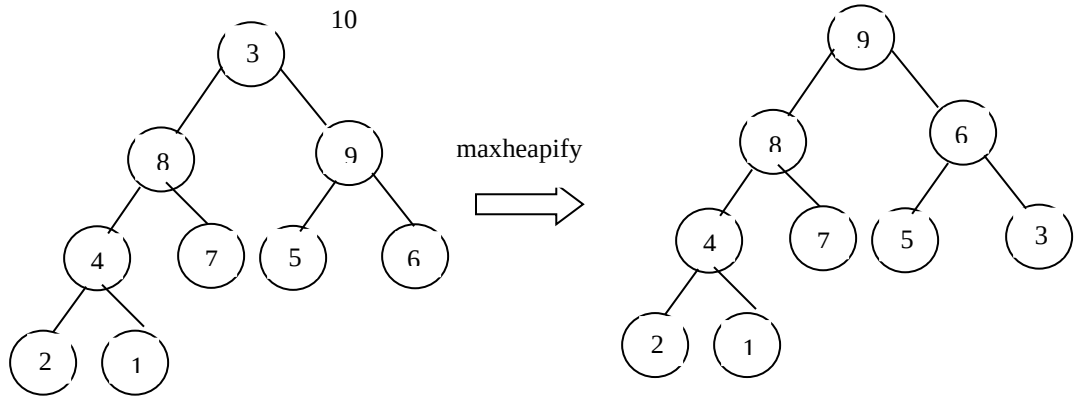
"RELATION_INDEXES" will have tuples describes each index. Each tuple will have attributes and their values. For the "WORKS_ON", there will be an index called I1 which is established on the combination of attributes SSN and PNO of the relation. The index name is "I1". The "MEMBER_ATTR" are SSN and PNO. There will be one tuple for each "MEMBER_ATTR" of this index in the "RELATION_INDEXES" table. The attribute number or key number for these the two attributes are 1 and therefore regarded as primary key of the relation WORKS_ON. The index type is "Primary" as it is established on primary keys of the relation. And it would be ascending (ASC) order in terms of its attributes values.

2. (30) Sort the following integer sequence:
1, 10, 9, 8, 7, 5, 6, 2, 4, 3
using the heap sorting algorithm. Trace the computation process.

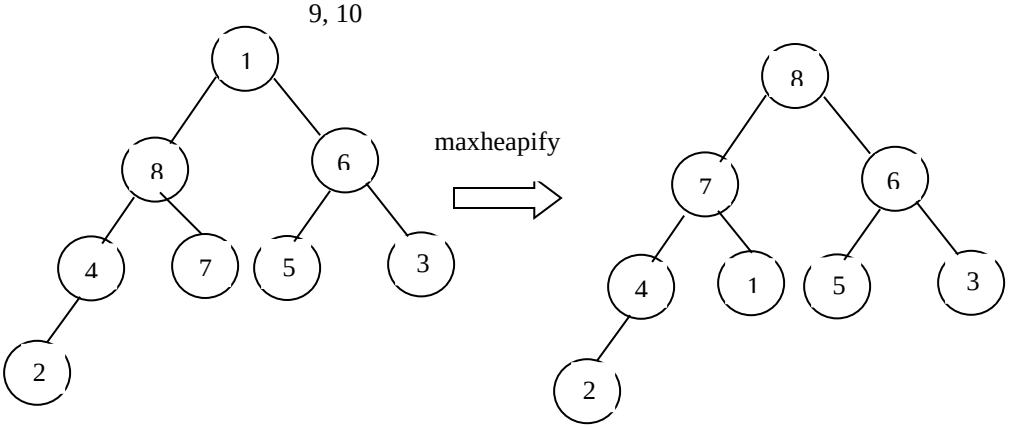
First step – construct a maxheap by the using buildMaxHeap algorithm (see in line 1Algorithm *heapSort* on page 32 in lecture note on Query Processing and Optimization):



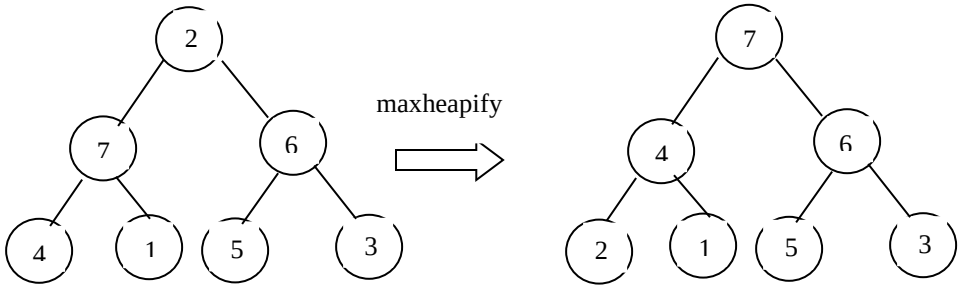
Second step: make a series of swapping-maxHeapify operations (see lines 2 -5 in Algorithm *heapSort* on page 32 in lecture note on Query Processing and Optimization):



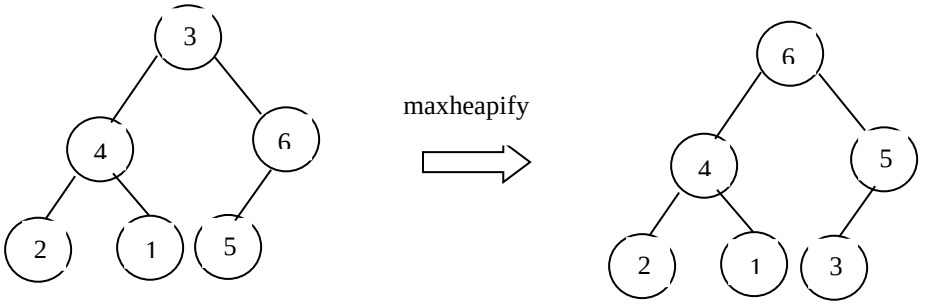
9, 10



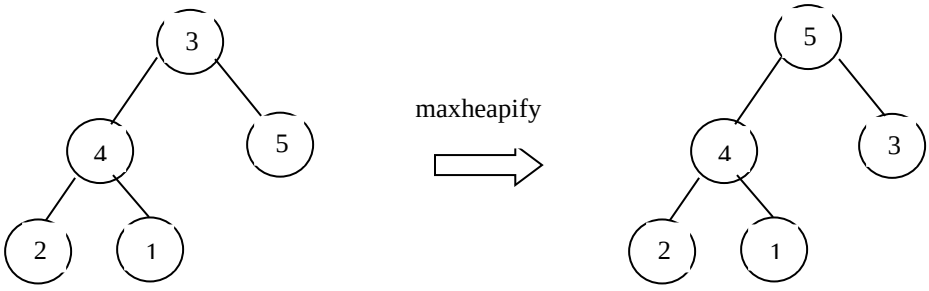
8, 9, 10



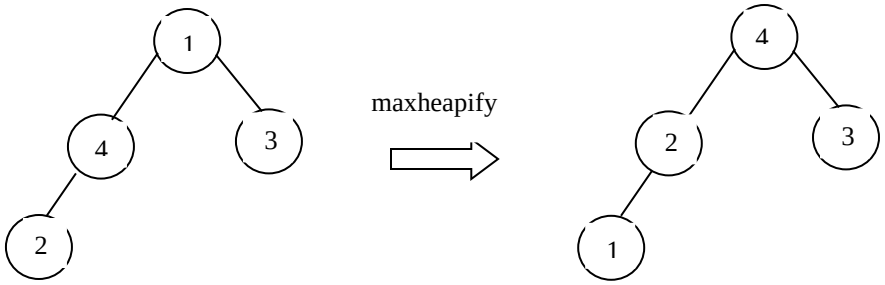
7, 8, 9, 10



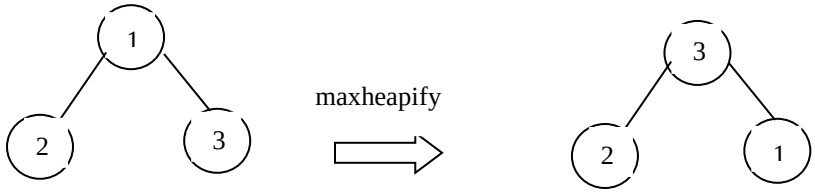
6, 7, 8, 9, 10



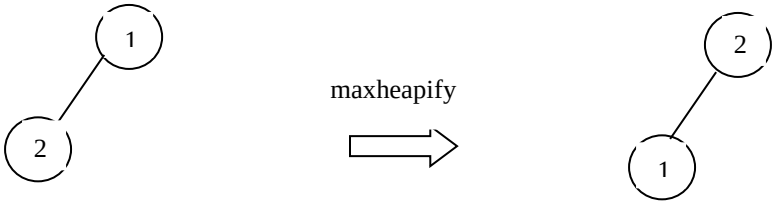
5, 6, 7, 8, 9, 10



4, 5, 6, 7, 8, 9, 10



3, 4, 5, 6, 7, 8, 9, 10



2, 3, 4, 5, 6, 7, 8, 9, 10

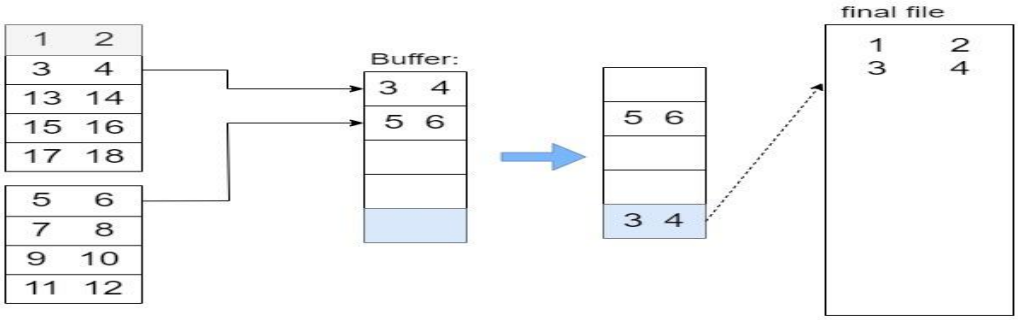
1, 2, 3, 4, 5, 6, 7, 8, 9, 10

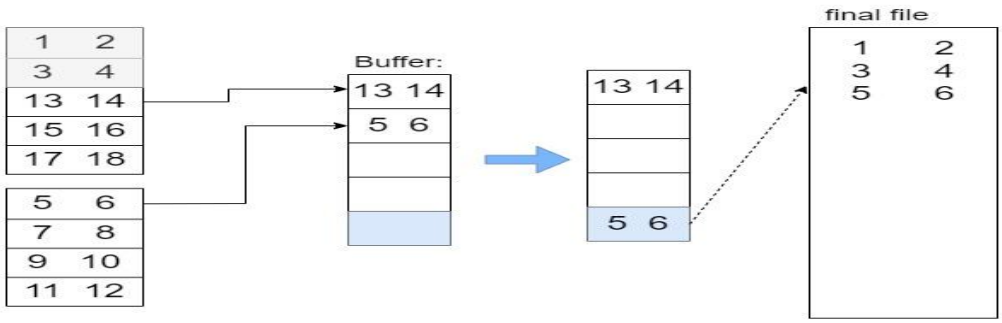
3. Assume that a file contains an integer sequence:

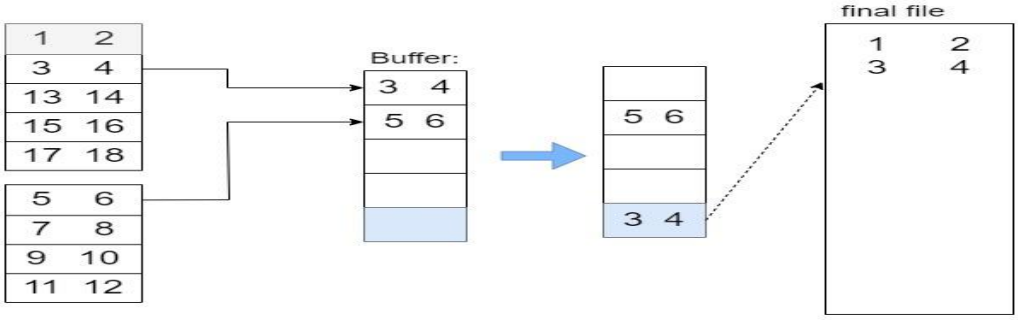
18, 17, 1, 2, 3, 4, 16, 15, 14, 13, 5, 6, 7, 8, 9, 10, 12, 11

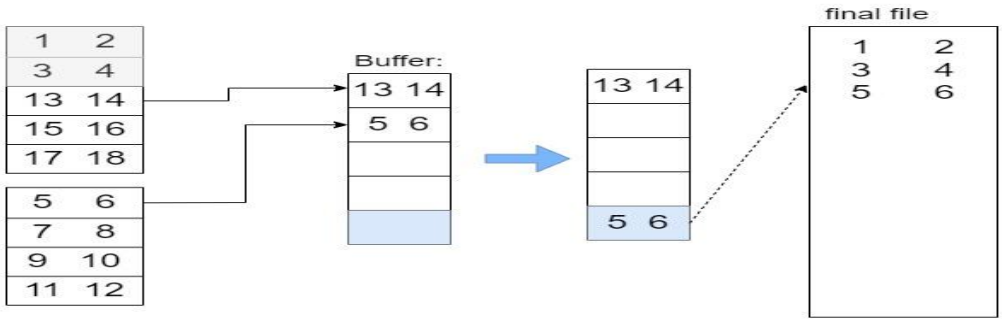
and the buffer (in main memory) is of size 5 blocks with each being able to hold 2 integers.

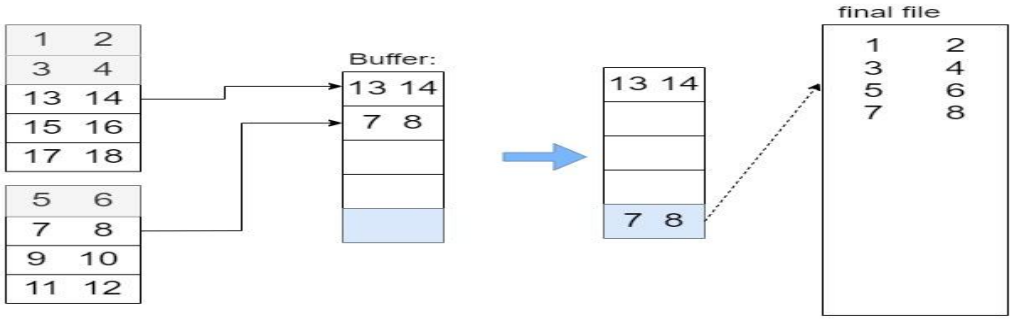
Sort this file using the external sorting algorithm. Trace the computation process.

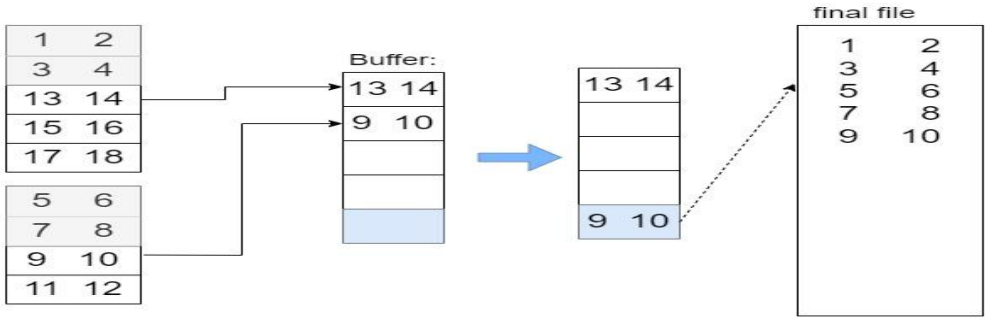

1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
3 4
5 6
5 6
3 4
final file
1 2
3 4

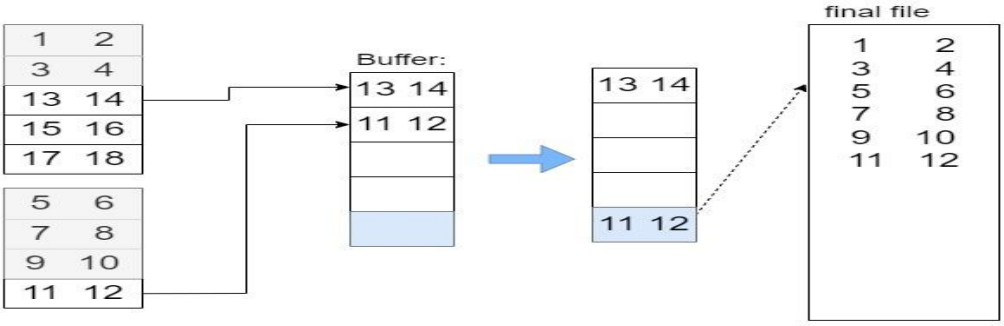

1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
13 14
5 6
13 14
5 6
final file
1 2
3 4
5 6

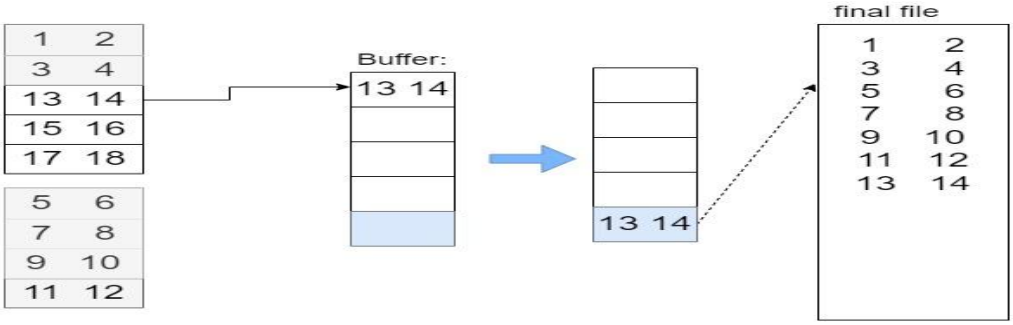

1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
3 4
5 6
5 6
3 4
final file
1 2
3 4


1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
13 14
5 6
13 14
5 6
final file
1 2
3 4
5 6


1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
13 14
7 8
13 14
7 8
final file
1 2
3 4
5 6
7 8


1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
13 14
9 10
13 14
9 10
final file
1 2
3 4
5 6
7 8
9 10


1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
13 14
11 12
13 14
11 12
final file
1 2
3 4
5 6
7 8
9 10
11 12


1 2
3 4
13 14
15 16
17 18
5 6
7 8
9 10
11 12
Buffer:
13 14
13 14
final file
1 2
3 4
5 6
7 8
9 10
11 12
13 14

4. (5) Given the following relation schema and assume that a primary index has been established on DNUMBER, how to evaluate $\sigma_{DNUMBER>5}$(DEPARTMENT) in an efficient way? Describe your method.

If a primary index has been established on DNUMBER, we can first find the record with DNUMBER = 5 in the file storing the DEPARTMENT table using the primary index. Then, return all the subsequent records in the file as the result.

5. (10) Given the following relation schema and assume that two indexes (including record points) have been established over superssn and dno, how to evaluate $\sigma_{dno = 4 \wedge superssn = 343488}$ (Employee) in an efficient way? Describe your method.

If we have two indexes over "dno" and "superssn", respectively, we can use them to get two subsets: $\{t_1, t_2, \ldots, t_n\}$ (employees working in dno = 4, and $\{s_1, s_2, \ldots, s_m\}$(employees supervised by ssn = 343488. The intersection of $\{t_1, t_2, \ldots, t_n\}$ and $\{s_1, s_2, \ldots, s_m\}$ will give the result, as illustrated in the following figure.