

DATA STORAGE, RETRIEVAL AND DATA BASE MANAGEMENT SYSTEMS

Question 1

- (a) Describe briefly, the following terms: (1 Mark Each)
- (i) Data Dictionary (PCC-May 2007 & Nov 2009)
 - (ii) Structured Query Language (PCC-May 2008)
 - (iii) DDL (PCC-Nov 2008)
 - (iv) DDL Compiler (PCC-June 2009)
- (b) Explain each of the following: (1 Mark Each)
- (i) Index Field (PCC-May 2007 & June 2009)
 - (ii) Data transformation (PCC-Nov 2007)
 - (iii) Transaction Log (PCC-May 2008 & Nov 2009)
 - (iv) Data Warehouse (PCC-Nov 2008)
 - (v) File Maintenance (PCC-Nov 2008)
 - (vi) Incremental Backup (PCC-Nov 2008)
 - (vii) Real Time Data Warehouse (PCC-June 2009)
 - (viii) Online Backup (PCC-June 2009)
 - (ix) Random Access (PCC-Nov 2009)

Answer

- (a) (i) Data Dictionary: A Data Dictionary is a set of metadata that contains definitions and representations of data elements. It maintains information pertaining to structure and usage of data and meta data.

Or

Data Dictionary: Data Dictionary maintains information pertaining to structure and usage of data and meta data. Each piece of data and various synonyms of data field are determined in consultation with database users.

Information Technology

- (ii) **Structured Query Language (SQL):** It is a set of thirty English like commands to create, update and access data from a database allowing users to raise adhoc queries interactively without the help of programmers.
 - (iii) **DDL :** DDL or Data Definition Language defines the conceptual schema providing a link between the logical and physical structures of the database. Some of the DDL functions define the physical characteristics of each record, field in the record, field's data type, its length, its logical name and also specify relationships among the records, describe the schema and subschema, provide means for associating related records or fields, provide for data security measures and logical and physical data independence.
 - (iv) **DDL Compiler:** DDL Compiler converts data definition statements into a set of tables. Tables contain meta-data (data about the data) concerning the database. It gives rise to a format that can be used by other components of the database.
- (b)
- (i) **Index field:** Index fields are used to store relevant information along with a documents. The data input to an index field is used to find those documents when needed. The program provides upto twenty five user definable index fields in an index set.
 - (ii) **Data transformation:** The data transformation layer receives data from the data sources, cleans and standardizes it, and loads it into data repository.
 - (iii) **Transaction Log:** A transaction log is a file that records database modifications such as insert, update, delete, commit, rollback and database schema changes. The database engine uses a transaction log to apply any changes made between the most recent checkpoint and the system failure.
 - (iv) **Data Warehouse:** It is a computer database that collects, integrates and stores an organization's data with the aim of producing accurate and timely management information and supporting data analysis. The data warehouses bring in data from a range of different data sources such as mainframe computers, micro computers as well as personal computers and office automation software such as spreadsheets and integrate this information in a single place.
 - (v) **File Maintenance:** File maintenance includes the updating of a file to reflect the effects of periodical changes by adding, altering data, eg. the addition of new programs to program library on magnetic disk.
 - (vi) **Incremental Backup:** This uses the DBBACKUP utility to copy the transaction log file since the most recent full backup. When incremental backup is performed, the mirror log is not backed up. When the user takes the backup and rename the log files, the transaction and mirror log file is renamed and new log files are created.
 - (vii) **Real time Data Warehouse:** A Real time data warehouse is updated on a transaction or event basis, every time an operational system performs a transaction such as an order or a delivery or a booking etc.

Data Storage, Retrieval and Data Base Management Systems

- (viii) Online Backup: Data base back-up can be performed while the database is being actively accessed (online). It is performed by executing the command-line or from the 'Backup Database' utility. When this process begins, the database engine externalizes all cached data pages kept in memory to the database file(s) on disk. This process is called a checkpoint. The database engine continues recording activity in the transaction log file while the database is backed up. The log file is backed up after the backup utility finishes backing up the database.
- (ix) Random Access: Random Access pertains to the method of file organization in a storage device in which the access time of the storage device is not significantly affected by the location of the data to be accessed. It means that any item of data which is stored online can be accessed within a relatively short time (usually in part of a second).

Question 2

Describe various factors which must be considered in determining the best file organisation for a particular application. (5 Marks) (PCC-May 2008 & Nov 2009)

Answer

Factors to be considered for best file organization are briefly discussed below:

- (i) File Volatility: It refers to the number of additions and deletions to the file in a given period of time. A file that constantly keeps changing is a highly volatile file. An Indexed-sequential file organization will not be suitable for such files, because additions have to be placed in the overflow area and constant reorganization of the file would have to occur. Other direct access methods would be a better choice. Even the sequential file organization could be appropriate if there are no interrogation requirements.
- (ii) File Activity: It is the proportion of master file records that are actually used or accessed in a given processing run. At one extreme is the real-time file where each transaction is processed immediately and hence at a time, only one master record is accessed. This situation obviously requires a direct access method. At the other extreme is a file, such as a payroll master file, where almost every record is accessed when the weekly payroll is processed. In such case, a sequentially ordered file would be more efficient.
- (iii) File Interrogation: It refers to the retrieval of information from a file. When the retrieval of individual record needs to be fast to support a real-time operation such as airline reservation, then some direct organization would be required. But if requirements of data can be delayed, then all the individual requests or information can be batched and run in a single processing run with a sequential file organization.
- (iv) File Size: Large files that require many individual references to records with immediate response, must be organized for certain direct access method. However, with small files, it may be more efficient to search sequentially or with more efficient binary search, to find an individual record.

Information Technology

Question 3

What are the management problems of file processing system?

(6 Marks & 3 Marks) (PCC-May 2007 & Nov 2008)

Answer

File processing in Information systems consist of using separate computer programs that update these independent data files and use them to produce the documents and reports required by each separate user application. However, many management problems are observed with File Processing Systems that limit the efficiency and effectiveness of end user applications:

- (i) **Data Duplication:** Independent data files include a lot of duplicated data that causes problems when data is to be updated, since separate file maintenance programs have to be developed and coordinated to ensure that each file is properly updated.
- (ii) **Lack of Data Integration:** Data in independent files makes it difficult to provide end users with information for ad hoc requests that require accessing data stored in several different files. Special computer programs have to be written to retrieve data from each independent file. This is difficult, time consuming, and expensive for the organizations.
- (iii) **Data Dependence:** In file processing systems, major components of a system i.e., the organization of files, their physical locations on storage, hardware and the application software used to access those files depend on one another in significant ways. Thus, if changes are made in the format and structure of data and records in a file, changes have to be made in all the programs that use this file. This program maintenance effort is a major burden of file processing systems.
- (iv) **Other Problems:** In file processing systems, data elements such as stock numbers and customer addresses are generally defined differently by different end users and applications. This causes serious inconsistency in the development of programs which access such data. In addition, the integrity (i.e. the accuracy and completeness) of the data is suspected because there is no control over their use and maintenance by authorized end users.

Question 4

What are the various views taken into account, while designing the architecture of a Database? Which view is user dependent and which one is user independent? Which view is storage device oriented?

(6 Marks)(PCC-June 2009)

Answer

The following three views are taken into account, while designing the architecture of a database.

- (i) External view (User View)
- (ii) Conceptual (Global view)
- (iii) Internal View (Physical view)

Data Storage, Retrieval and Data Base Management Systems

External view (User View) encircles the following:

- ◆ It is at the highest level of the database abstraction.
- ◆ It includes only those portions of database or application programs which are of concern to the users.
- ◆ It is described by means of a scheme, called the external schema.
- ◆ It is defined by the users or written by the programmers.

Conceptual (Global view) which is viewed by the Data Base Administrator, encompasses the following –

- ◆ All database entities and relationships among them are included.
- ◆ Single view represents the entire database.
- ◆ It is defined by the conceptual schema.
- ◆ It describes all records, relationships and constraints or boundaries.
- ◆ Data description to render it independent of the physical representation.

Internal View (Physical view) contains the following:

- ◆ It is at the lowest level of database abstraction.
- ◆ It is closest to the physical storage method.
- ◆ It indicates how data will be stored.
- ◆ It describes data structure.
- ◆ It describes access methods.
- ◆ It is expressed by internal schema.

External view is user-dependent as external view is also referred as User View.

Conceptual and Internal views are user-independent.

Internal view is storage device oriented.

Question 5

Discuss salient features of Hierarchical Database structure. (5 Marks)(PCC-Nov 2007)

Answer

Salient features of Hierarchical Database:

In a hierarchical database, records are logically organized into hierarchy of relationship. It is arranged in an inverted tree pattern. The following are salient features:

- ❖ Database structure is less flexible as relationship between records are relatively fixed by the structure.
- ❖ It requires that hierarchy of records must be determined and implemented before a search.

Information Technology

- ❖ Ad hoc queries are difficult and time consuming to accomplish.
- ❖ Frequent management queries may not be supported as effectively.
- ❖ Day to day operational data can be processed rapidly.
- ❖ Any group of records with natural relation may fit nicely.
- ❖ Records are logically structured in inverted tree pattern.
- ❖ It provides the parent – child relationship amongst the nodes.
- ❖ It implements one-to-one and one-to-many relationship.

Question 6

Describe DDL and DML.

(5 Marks) (PCC-May 2007)

Answer

- (i) Data Definition Language (DDL): It defines the conceptual schema providing a link between the logical and physical structures of the database. The logical structure of a database is a schema. A subschema is the way a specific application views the data from the database.

Following are the functions of Data Definition Language (DDL):

- (a) They define the physical characteristics of each record, field in the record, field's type and length, field's logical name and also specify relationships among the records.
- (b) They describe the schema and subschema.
- (c) They indicate the keys of the record.
- (d) They provide means for associating related records or fields.
- (e) They provide for data security measures.
- (f) They provide for logical and physical data independence.

- (ii) Data Manipulation Language (DML): DML is a Database Language used by data base users to retrieve, insert, delete and update data in a database.

Following are the functions of Data Manipulation Language (DML):

- (a) They provide the data manipulation techniques like deletion, modification, insertion, replacement, retrieval, sorting and display of data or records.
- (b) They facilitate use of relationships between the records.
- (c) They enable the user and application program to be independent of the physical data structures and database structures maintenance by allowing to process data on a logical and symbolic basis rather than on a physical location basis.
- (d) They provide for independence of programming languages by supporting several high-level procedural languages like COBOL, PL/1 and C++.

Question 7

Data Storage, Retrieval and Data Base Management Systems

Why documentation is required? List any 4 types of documentations required to be prepared prior to delivery of customized software to a customer. (6 Marks) (PCC-June 2009)

Answer

The documentation is an important aspect of Software Development Life Cycle which provides a method to understand the various issues related with software development and provide a method to access details related to system study, system development, system testing, system operational details, details related to preventive maintenance and details associated with further modification aspects of the software.

Four important documentations required to be prepared prior to delivery of customized software to customer are as follows:

- (i) Strategic and Application Plans.
- (ii) Application Systems and Program Documentation.
- (iii) System Software and Utility Program Documentation.
- (iv) Database Documentation, Operation Manuals, User Manuals, Testing Manual, Standard Manual, Preventive Maintenance Manual, and Backup Manual are other important documentations.

Question 8

Write short notes on the following:

- (i) Data warehouse (5 Marks)(PCC-May 2008)

Answer

- (i) Data warehouse: It is a computer database that collects, integrates and stores an organization's data with the aim of producing accurate and timely management information and supporting data analysis. It was developed to meet a growing demand for management information and analysis that could not be met by operational systems. Operational systems were unable to meet this need for a range of reasons:

- ◆ The processing load of reporting reduced the response time of the operational systems.
- ◆ The database designs of operational systems were not optimised for information analysis and reporting.
- ◆ Most organizations had more than one operational system, so company-wide reporting could not be supported from a single system.
- ◆ Development of reports in operational systems often requires writing specific computer programs, which is slow and expensive.

As a result, separate computer databases began to be built that were specifically designed to support management information and analysis purposes. These data warehouses were able to bring in data from a range of different data sources.

Information Technology

The advantages of using a data warehouse are:

- ◆ Enhanced end-user access to a wide variety of data.
- ◆ Increased data consistency.
- ◆ Increased productivity and decreases computing costs.
- ◆ It is able to combine data from different sources, in one place.
- ◆ It provides an infrastructure that could support changes to data and replication of the changed data back into the operational systems.